Wrangle and Analyse Data¶

In the following project I am going to gather and analyze data all around the Twitter account WeRateDogs. For that, data is gathered via manual download, programmatical download and over an API. After that I am going to assess this data, define the issues found during the assessment and clean these issues to get a cleaned master dataframe.

Libraries needed¶

In [0]:

```
import requests #download data
import numpy as np #array functions
import pandas as pd #data handling
import tweepy #twitter api
import json #handle json data
import matplotlib.pyplot as plt #data visualization
import seaborn as sns #data visualization
import re #text processing
```

```
/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning:
pandas.util.testing is deprecated. Use the functions in the public API at
pandas.testing instead.
  import pandas.util.testing as tm
```

Read the csv file

In [0]:

```
url = 'https://raw.githubusercontent.com/hiyabose/Data-Analysis/master/twitter-archive-
enhanced.csv'
df_twitter = pd.read_csv(url)
```

```
df_twitter.head(3)
```

Out[0]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	retweeted_status_e
0	892420643555336193	NaN	NaN	2017-08-01 16:23:56 +0000	<a href="http://twitter.com/download/iphon	This is Phineas. He's a mystical	NaN	NaN	NaN

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	retweeted_sta...e
					e" r...	boy. Only eve...			
1	892177421306343426	NaN	NaN	2017-08-01 00:17:27 +0000	<a href="http://twitter.com/download/iphone" r...	This is Tilly. She's just checking pup on you....	NaN	NaN	NaN
2	891815181378084864	NaN	NaN	2017-07-31 00:18:03 +0000	<a href="http://twitter.com/download/iphone" r...	This is Archie. He is a rare Norwegian Pouncin...	NaN	NaN	NaN

Tweet image predictions (image_predictions.tsv)

To gather this data we are going to define the file - url, request this url and write the content of the response to a separate file.

In [0]:

```
url = "https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv"

#get response
response = requests.get(url)

#write return to an image
with open("image_predictions.tsv", mode = "wb") as file:
    file.write(response.content)
```

In [0]:

```
df_predict = pd.read_csv("image_predictions.tsv", sep='\t')

df_predict.head(3)
```

Out[0]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog
0	666020888028790149	https://pbs.twimg.com/media/	1	Welsh_springer_sp	0.465074	True	collie	0.156665	True	Sh...ee

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	
		CT4udn0WwAA0aMy.jpg		aniel						
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	1	redbone	0.506826	True	miniature_pinscher	0.074192	True	Rh ric
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	1	German_shepherd	0.596461	True	malinois	0.138584	True	bl

In [0]:

```
#keys and token to access the API
consumer_key = 'vhL3DZrG6B7uqGJzPjve24yID'
consumer_secret = 'mh0iZy07tCB6wFR6kv1kEZcEqH7oz5agoYjTDDFSl1DklajImU'
access_token = '1257573345849114625-SSDZM4Z40M6Dk36Y6x1sj7mWMomp3I'
access_secret = 'Y3yGlzJFbNWdkNPazro3GYwvq2pKl41nh5Q4DRxOUsN3E'

#access the API
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)

api = tweepy.API(auth, wait_on_rate_limit=True, wait_on_rate_limit_notify=True)
```

In [0]:

```
twitter_ids = list(df_twitter.tweet_id.unique())
```

In [0]:

```
with open("tweet_json.txt", "w") as file:
  for ids in twitter_ids:
    print(f"Gather id: {ids}")
    try:

        #get all the twitter status - extended mode gives us additional data
        tweet = api.get_status(ids, tweet_mode = "extended")
        #dump the json data to our file
        json.dump(tweet._json, file)
        #add a linebreak after each dump
        file.write('\n')
    except Exception as e:
        print(f"Error - id: {ids}" + str(e))
```

```
Gather id: 892420643555336193
Gather id: 892177421306343426
Gather id: 891815181378084864
```

Gather id: 891689557279858688
Gather id: 891327558926688256
Gather id: 891087950875897856
Gather id: 890971913173991426
Gather id: 890729181411237888
Gather id: 890609185150312448
Gather id: 890240255349198849
Gather id: 890006608113172480
Gather id: 889880896479866881
Gather id: 889665388333682689
Gather id: 889638837579907072
Gather id: 889531135344209921
Gather id: 889278841981685760
Gather id: 888917238123831296
Gather id: 888804989199671297
Gather id: 888554962724278272
Gather id: 888202515573088257
Error - id: 888202515573088257[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 888078434458587136
Gather id: 887705289381826560
Gather id: 887517139158093824
Gather id: 887473957103951883
Gather id: 887343217045368832
Gather id: 887101392804085760
Gather id: 886983233522544640
Gather id: 886736880519319552
Gather id: 886680336477933568
Gather id: 886366144734445568
Gather id: 886267009285017600
Gather id: 886258384151887873
Gather id: 886054160059072513
Gather id: 885984800019947520
Gather id: 885528943205470208
Gather id: 885518971528720385
Gather id: 885311592912609280
Gather id: 885167619883638784
Gather id: 884925521741709313
Gather id: 884876753390489601
Gather id: 884562892145688576
Gather id: 884441805382717440
Gather id: 884247878851493888
Gather id: 884162670584377345
Gather id: 883838122936631299

Gather id: 883482846933004288
Gather id: 883360690899218434
Gather id: 883117836046086144
Gather id: 882992080364220416
Gather id: 882762694511734784
Gather id: 882627270321602560
Gather id: 882268110199369728
Gather id: 882045870035918850
Gather id: 881906580714921986
Gather id: 881666595344535552
Gather id: 881633300179243008
Gather id: 881536004380872706
Gather id: 881268444196462592
Gather id: 880935762899988482
Gather id: 880872448815771648
Gather id: 880465832366813184
Gather id: 880221127280381952
Gather id: 880095782870896641
Gather id: 879862464715927552
Gather id: 879674319642796034
Gather id: 879492040517615616
Gather id: 879415818425184262
Gather id: 879376492567855104
Gather id: 879130579576475649
Gather id: 879050749262655488
Gather id: 879008229531029506
Gather id: 878776093423087618
Gather id: 878604707211726852
Gather id: 878404777348136964
Gather id: 878316110768087041
Gather id: 878281511006478336
Gather id: 878057613040115712
Gather id: 877736472329191424
Gather id: 877611172832227328
Gather id: 877556246731214848
Gather id: 877316821321428993
Gather id: 877201837425926144
Gather id: 876838120628539392
Gather id: 876537666061221889
Gather id: 876484053909872640
Gather id: 876120275196170240
Gather id: 875747767867523072
Gather id: 875144289856114688
Gather id: 875097192612077568

```
Gather id: 875021211251597312
Gather id: 874680097055178752
Gather id: 874434818259525634
Gather id: 874296783580663808
Gather id: 874057562936811520
Gather id: 874012996292530176
Gather id: 873697596434513921
Error - id: 873697596434513921[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 873580283840344065
Gather id: 873337748698140672
Gather id: 873213775632977920
Gather id: 872967104147763200
Gather id: 872820683541237760
Gather id: 872668790621863937
Error - id: 872668790621863937[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 872620804844003328
Gather id: 872486979161796608
Gather id: 872261713294495745
Error - id: 872261713294495745[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 872122724285648897
Gather id: 871879754684805121
Gather id: 871762521631449091
Gather id: 871515927908634625
Gather id: 871166179821445120
Gather id: 871102520638267392
Gather id: 871032628920680449
Gather id: 870804317367881728
Gather id: 870726314365509632
Gather id: 870656317836468226
Gather id: 870374049280663552
Gather id: 870308999962521604
Gather id: 870063196459192321
Gather id: 869988702071779329
Error - id: 869988702071779329[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 869772420881756160
Gather id: 869702957897576449
Gather id: 869596645499047938
Gather id: 869227993411051520
Gather id: 868880397819494401
Gather id: 868639477480148993
```

```
Gather id: 868622495443632128
Gather id: 868552278524837888
Gather id: 867900495410671616
Gather id: 867774946302451713
Gather id: 867421006826221569
Gather id: 867072653475098625
Gather id: 867051520902168576
Gather id: 866816280283807744
Error - id: 866816280283807744[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 866720684873056260
Gather id: 866686824827068416
Gather id: 866450705531457537
Gather id: 866334964761202691
Gather id: 866094527597207552
Gather id: 865718153858494464
Gather id: 865359393868664832
Gather id: 865006731092295680
Gather id: 864873206498414592
Gather id: 864279568663928832
Gather id: 864197398364647424
Gather id: 863907417377173506
Gather id: 863553081350529029
Gather id: 863471782782697472
Gather id: 863432100342583297
Gather id: 863427515083354112
Gather id: 863079547188785154
Gather id: 863062471531167744
Gather id: 862831371563274240
Gather id: 862722525377298433
Gather id: 862457590147678208
Gather id: 862096992088072192
Gather id: 861769973181624320
Error - id: 861769973181624320[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 861383897657036800
Gather id: 861288531465048066
Gather id: 861005113778896900
Gather id: 860981674716409858
Gather id: 860924035999428608
Gather id: 860563773140209665
Gather id: 860524505164394496
Gather id: 860276583193509888
Gather id: 860184849394610176
```

```
Gather id: 860177593139703809
Gather id: 859924526012018688
Gather id: 859851578198683649
Gather id: 859607811541651456
Gather id: 859196978902773760
Gather id: 859074603037188101
Gather id: 858860390427611136
Gather id: 858843525470990336
Gather id: 858471635011153920
Gather id: 858107933456039936
Gather id: 857989990357356544
Gather id: 857746408056729600
Gather id: 857393404942143489
Gather id: 857263160327368704
Gather id: 857214891891077121
Gather id: 857062103051644929
Gather id: 857029823797047296
Gather id: 856602993587888130
Error - id: 856602993587888130[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 856543823941562368
Gather id: 856526610513747968
Gather id: 856330835276025856
Gather id: 856288084350160898
Gather id: 856282028240666624
Gather id: 855862651834028034
Gather id: 855860136149123072
Gather id: 855857698524602368
Gather id: 855851453814013952
Gather id: 855818117272018944
Gather id: 855459453768019968
Gather id: 855245323840757760
Gather id: 855138241867124737
Gather id: 854732716440526848
Gather id: 854482394044301312
Gather id: 854365224396361728
Gather id: 854120357044912130
Gather id: 854010172552949760
Gather id: 853760880890318849
Gather id: 853639147608842240
Gather id: 853299958564483072
Gather id: 852936405516943360
Gather id: 852912242202992640
Gather id: 852672615818899456
```

```
Gather id: 852553447878664193
Gather id: 852311364735569921
Gather id: 852226086759018497
Gather id: 852189679701164033
Gather id: 851953902622658560
Error - id: 851953902622658560[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 851861385021730816
Gather id: 851591660324737024
Gather id: 851464819735769094
Gather id: 851224888060895234
Gather id: 850753642995093505
Gather id: 850380195714523136
Gather id: 850333567704068097
Gather id: 850145622816686080
Gather id: 850019790995546112
Gather id: 849776966551130114
Gather id: 849668094696017920
Gather id: 849412302885593088
Gather id: 849336543269576704
Gather id: 849051919805034497
Gather id: 848690551926992896
Gather id: 848324959059550208
Gather id: 848213670039564288
Gather id: 848212111729840128
Gather id: 847978865427394560
Gather id: 847971574464610304
Gather id: 847962785489326080
Gather id: 847842811428974592
Gather id: 847617282490613760
Gather id: 847606175596138505
Gather id: 847251039262605312
Gather id: 847157206088847362
Gather id: 847116187444137987
Gather id: 846874817362120707
Gather id: 846514051647705089
Gather id: 846505985330044928
Gather id: 846153765933735936
Gather id: 846139713627017216
Gather id: 846042936437604353
Gather id: 845812042753855489
Gather id: 845677943972139009
Gather id: 845459076796616705
Error - id: 845459076796616705[{'code': 144, 'message': 'No status found with that
```

ID.'}]
Gather id: 845397057150107648
Gather id: 845306882940190720
Gather id: 845098359547420673
Gather id: 844979544864018432
Gather id: 844973813909606400
Gather id: 844704788403113984
Error - id: 844704788403113984[{'code': 144, 'message': 'No status found with that ID.'}]
Gather id: 844580511645339650
Gather id: 844223788422217728
Gather id: 843981021012017153
Gather id: 843856843873095681
Gather id: 843604394117681152
Gather id: 843235543001513987
Gather id: 842892208864923648
Error - id: 842892208864923648[{'code': 144, 'message': 'No status found with that ID.'}]
Gather id: 842846295480000512
Gather id: 842765311967449089
Gather id: 842535590457499648
Gather id: 842163532590374912
Gather id: 842115215311396866
Gather id: 841833993020538882
Gather id: 841680585030541313
Gather id: 841439858740625411
Gather id: 841320156043304961
Gather id: 841314665196081154
Gather id: 841077006473256960
Gather id: 840761248237133825
Gather id: 840728873075638272
Gather id: 840698636975636481
Gather id: 840696689258311684
Gather id: 840632337062862849
Gather id: 840370681858686976
Gather id: 840268004936019968
Gather id: 839990271299457024
Gather id: 839549326359670784
Gather id: 839290600511926273
Gather id: 839239871831150596
Gather id: 838952994649550848
Gather id: 838921590096166913
Gather id: 838916489579200512
Gather id: 838831947270979586

Gather id: 838561493054533637
Gather id: 838476387338051585
Gather id: 838201503651401729
Gather id: 838150277551247360
Gather id: 838085839343206401
Gather id: 838083903487373313
Gather id: 837820167694528512
Gather id: 837482249356513284
Gather id: 837471256429613056
Gather id: 837366284874571778
Error - id: 837366284874571778[{'code': 144, 'message': 'No status found with that ID.'}]
Gather id: 837110210464448512
Gather id: 837012587749474308
Error - id: 837012587749474308[{'code': 144, 'message': 'No status found with that ID.'}]
Gather id: 836989968035819520
Gather id: 836753516572119041
Gather id: 836677758902222849
Gather id: 836648853927522308
Gather id: 836397794269200385
Gather id: 836380477523124226
Gather id: 836260088725786625
Gather id: 836001077879255040
Gather id: 835685285446955009
Gather id: 835574547218894849
Gather id: 835536468978302976
Gather id: 835309094223372289
Gather id: 835297930240217089
Gather id: 835264098648616962
Gather id: 835246439529840640
Gather id: 835172783151792128
Gather id: 835152434251116546
Gather id: 834931633769889797
Gather id: 834786237630337024
Gather id: 834574053763584002
Gather id: 834477809192075265
Gather id: 834458053273591808
Gather id: 834209720923721728
Gather id: 834167344700198914
Gather id: 834089966724603904
Gather id: 834086379323871233
Gather id: 833863086058651648
Gather id: 833826103416520705

```
Gather id: 833732339549220864
Gather id: 833722901757046785
Gather id: 833479644947025920
Gather id: 833124694597443584
Gather id: 832998151111966721
Gather id: 832769181346996225
Gather id: 832757312314028032
Gather id: 832682457690300417
Gather id: 832645525019123713
Gather id: 832636094638288896
Gather id: 832397543355072512
Gather id: 832369877331693569
Gather id: 832273440279240704
Gather id: 832215909146226688
Gather id: 832215726631055365
Gather id: 832088576586297345
Gather id: 832040443403784192
Gather id: 832032802820481025
Gather id: 831939777352105988
Gather id: 831926988323639298
Gather id: 831911600680497154
Gather id: 831670449226514432
Gather id: 831650051525054464
Gather id: 831552930092285952
Gather id: 831322785565769729
Gather id: 831315979191906304
Gather id: 831309418084069378
Gather id: 831262627380748289
Gather id: 830956169170665475
Gather id: 830583320585068544
Gather id: 830173239259324417
Gather id: 830097400375152640
Gather id: 829878982036299777
Gather id: 829861396166877184
Gather id: 829501995190984704
Gather id: 829449946868879360
Gather id: 829374341691346946
Error - id: 829374341691346946[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 829141528400556032
Gather id: 829011960981237760
Gather id: 828801551087042563
Gather id: 828770345708580865
Gather id: 828708714936930305
```

```
Gather id: 828650029636317184
Gather id: 828409743546925057
Gather id: 828408677031882754
Gather id: 828381636999917570
Gather id: 828376505180889089
Gather id: 828372645993398273
Gather id: 828361771580813312
Gather id: 828046555563323392
Gather id: 828011680017821696
Gather id: 827933404142436356
Gather id: 827653905312006145
Gather id: 827600520311402496
Gather id: 827324948884643840
Gather id: 827228250799742977
Error - id: 827228250799742977[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 827199976799354881
Gather id: 826958653328592898
Gather id: 826848821049180160
Gather id: 826615380357632002
Gather id: 826598799820865537
Gather id: 826598365270007810
Gather id: 826476773533745153
Gather id: 826240494070030336
Gather id: 826204788643753985
Gather id: 826115272272650244
Gather id: 825876512159186944
Gather id: 825829644528148480
Gather id: 825535076884762624
Gather id: 825147591692263424
Gather id: 825120256414846976
Gather id: 825026590719483904
Gather id: 824796380199809024
Gather id: 824775126675836928
Gather id: 824663926340194305
Gather id: 824325613288833024
Gather id: 824297048279236611
Gather id: 824025158776213504
Gather id: 823939628516474880
Gather id: 823719002937630720
Gather id: 823699002998870016
Gather id: 823581115634085888
Gather id: 823333489516937216
Gather id: 823322678127919110
```

```
Gather id: 823269594223824897
Gather id: 822975315408461824
Gather id: 822872901745569793
Gather id: 822859134160621569
Gather id: 822647212903690241
Gather id: 822610361945911296
Gather id: 822489057087389700
Gather id: 822462944365645825
Gather id: 822244816520155136
Gather id: 822163064745328640
Gather id: 821886076407029760
Gather id: 821813639212650496
Gather id: 821765923262631936
Gather id: 821522889702862852
Gather id: 821421320206483457
Gather id: 821407182352777218
Gather id: 821153421864615936
Gather id: 821149554670182400
Gather id: 821107785811234820
Gather id: 821044531881721856
Gather id: 820837357901512704
Gather id: 820749716845686786
Gather id: 820690176645140481
Gather id: 820494788566847489
Gather id: 820446719150292993
Gather id: 820314633777061888
Gather id: 820078625395449857
Gather id: 820013781606658049
Gather id: 819952236453363712
Gather id: 819924195358416896
Gather id: 819711362133872643
Gather id: 819588359383371776
Gather id: 819347104292290561
Gather id: 819238181065359361
Gather id: 819227688460238848
Gather id: 819015337530290176
Gather id: 819015331746349057
Gather id: 819006400881917954
Gather id: 819004803107983360
Gather id: 818646164899774465
Gather id: 818627210458333184
Gather id: 818614493328580609
Gather id: 818588835076603904
Gather id: 818536468981415936
```

```
Gather id: 818307523543449600
Gather id: 818259473185828864
Gather id: 818145370475810820
Gather id: 817908911860748288
Gather id: 817827839487737858
Gather id: 817777686764523521
Gather id: 817536400337801217
Gather id: 817502432452313088
Gather id: 817423860136083457
Gather id: 817415592588222464
Gather id: 817181837579653120
Gather id: 817171292965273600
Gather id: 817120970343411712
Gather id: 817056546584727552
Gather id: 816829038950027264
Gather id: 816816676327063552
Gather id: 816697700272001025
Gather id: 816450570814898180
Gather id: 816336735214911488
Gather id: 816091915477250048
Gather id: 816062466425819140
Gather id: 816014286006976512
Gather id: 815990720817401858
Gather id: 815966073409433600
Gather id: 815745968457060357
Gather id: 815736392542261248
Gather id: 815639385530101762
Gather id: 815390420867969024
Gather id: 814986499976527872
Gather id: 814638523311648768
Gather id: 814578408554463233
Gather id: 814530161257443328
Gather id: 814153002265309185
Gather id: 813944609378369540
Gather id: 813910438903693312
Gather id: 813812741911748608
Gather id: 813800681631023104
Gather id: 813217897535406080
Gather id: 813202720496779264
Gather id: 813187593374461952
Gather id: 813172488309972993
Gather id: 813157409116065792
Gather id: 813142292504645637
Gather id: 813130366689148928
```

Gather id: 813127251579564032
Gather id: 813112105746448384
Gather id: 813096984823349248
Gather id: 813081950185472002
Gather id: 813066809284972545
Gather id: 813051746834595840
Gather id: 812781120811126785
Gather id: 812747805718642688
Error - id: 812747805718642688[{'code': 144, 'message': 'No status found with that ID.'}]
Gather id: 812709060537683968
Gather id: 812503143955202048
Gather id: 812466873996607488
Gather id: 812372279581671427
Gather id: 811985624773361665
Gather id: 811744202451197953
Gather id: 811647686436880384
Gather id: 811627233043480576
Gather id: 811386762094317568
Gather id: 810984652412424192
Gather id: 810896069567610880
Gather id: 810657578271330305
Gather id: 810284430598270976
Gather id: 810254108431155201
Gather id: 809920764300447744
Gather id: 809808892968534016
Gather id: 809448704142938112
Gather id: 809220051211603969
Gather id: 809084759137812480
Gather id: 808838249661788160
Gather id: 808733504066486276
Gather id: 808501579447930884
Gather id: 808344865868283904
Gather id: 808134635716833280
Gather id: 808106460588765185
Gather id: 808001312164028416
Gather id: 807621403335917568
Gather id: 807106840509214720
Gather id: 807059379405148160
Gather id: 807010152071229440
Gather id: 806629075125202948
Gather id: 806620845233815552
Gather id: 806576416489959424
Gather id: 806542213899489280

Gather id: 806242860592926720
Gather id: 806219024703037440
Gather id: 805958939288408065
Gather id: 805932879469572096
Gather id: 805826884734976000
Gather id: 805823200554876929
Gather id: 805520635690676224
Gather id: 805487436403003392
Gather id: 805207613751304193
Gather id: 804738756058218496
Gather id: 804475857670639616
Gather id: 804413760345620481
Gather id: 804026241225523202
Gather id: 803773340896923648
Gather id: 803692223237865472
Gather id: 803638050916102144
Gather id: 803380650405482500
Gather id: 803321560782307329
Gather id: 803276597545603072
Gather id: 802952499103731712
Gather id: 802624713319034886
Gather id: 802600418706604034
Gather id: 802572683846291456
Gather id: 802323869084381190
Gather id: 802265048156610565
Gather id: 802247111496568832
Error - id: 802247111496568832[{'code': 144, 'message': 'No status found with that ID.'}]
Gather id: 802239329049477120
Gather id: 802185808107208704
Gather id: 801958328846974976
Gather id: 801854953262350336
Gather id: 801538201127157760
Gather id: 801285448605831168
Gather id: 801167903437357056
Gather id: 801127390143516673
Gather id: 801115127852503040
Gather id: 800859414831898624
Gather id: 800855607700029440
Gather id: 800751577355128832
Gather id: 800513324630806528
Gather id: 800459316964663297
Gather id: 800443802682937345
Gather id: 800388270626521089

Gather id: 800188575492947969
Gather id: 800141422401830912
Gather id: 800018252395122689
Gather id: 799774291445383169
Gather id: 799757965289017345
Gather id: 799422933579902976
Gather id: 799308762079035393
Gather id: 799297110730567681
Gather id: 799063482566066176
Gather id: 798933969379225600
Gather id: 798925684722855936
Gather id: 798705661114773508
Gather id: 798701998996647937
Gather id: 798697898615730177
Gather id: 798694562394996736
Gather id: 798686750113755136
Gather id: 798682547630837760
Gather id: 798673117451325440
Gather id: 798665375516884993
Gather id: 798644042770751489
Gather id: 798628517273620480
Gather id: 798585098161549313
Gather id: 798576900688019456
Gather id: 798340744599797760
Gather id: 798209839306514432
Gather id: 797971864723324932
Gather id: 797545162159308800
Gather id: 797236660651966464
Gather id: 797165961484890113
Gather id: 796904159865868288
Gather id: 796865951799083009
Gather id: 796759840936919040
Gather id: 796563435802726400
Gather id: 796484825502875648
Gather id: 796387464403357696
Gather id: 796177847564038144
Gather id: 796149749086875649
Gather id: 796125600683540480
Gather id: 796116448414461957
Gather id: 796080075804475393
Gather id: 796031486298386433
Gather id: 795464331001561088
Gather id: 795400264262053889
Gather id: 795076730285391872

```
Gather id: 794983741416415232
Gather id: 794926597468000259
Gather id: 794355576146903043
Gather id: 794332329137291264
Gather id: 794205286408003585
Gather id: 793962221541933056
Gather id: 793845145112371200
Gather id: 793614319594401792
Gather id: 793601777308463104
Gather id: 793500921481273345
Gather id: 793286476301799424
Gather id: 793271401113350145
Gather id: 793256262322548741
Gather id: 793241302385262592
Gather id: 793226087023144960
Gather id: 793210959003287553
Gather id: 793195938047070209
Gather id: 793180763617361921
Gather id: 793165685325201412
Gather id: 793150605191548928
Gather id: 793135492858580992
Gather id: 793120401413079041
Gather id: 792913359805018113
Gather id: 792883833364439040
Gather id: 792773781206999040
Gather id: 792394556390137856
Gather id: 792050063153438720
Gather id: 791821351946420224
Gather id: 791784077045166082
Gather id: 791780927877898241
Gather id: 791774931465953280
Gather id: 791672322847637504
Gather id: 791406955684368384
Gather id: 791312159183634433
Gather id: 791026214425268224
Gather id: 790987426131050500
Gather id: 790946055508652032
Gather id: 790723298204217344
Gather id: 790698755171364864
Gather id: 790581949425475584
Gather id: 790337589677002753
Gather id: 790277117346975746
Gather id: 790227638568808452
Gather id: 789986466051088384
```

```
Gather id: 789960241177853952
Gather id: 789903600034189313
Gather id: 789628658055020548
Gather id: 789599242079838210
Gather id: 789530877013393408
Gather id: 789314372632018944
Gather id: 789280767834746880
Gather id: 789268448748703744
Gather id: 789137962068021249
Gather id: 788908386943430656
Gather id: 788765914992902144
Gather id: 788552643979468800
Gather id: 788412144018661376
Gather id: 788178268662984705
Gather id: 788150585577050112
Gather id: 788070120937619456
Gather id: 788039637453406209
Gather id: 787810552592695296
Gather id: 787717603741622272
Gather id: 787397959788929025
Gather id: 787322443945877504
Gather id: 787111942498508800
Gather id: 786963064373534720
Gather id: 786729988674449408
Gather id: 786709082849828864
Gather id: 786664955043049472
Gather id: 786595970293370880
Gather id: 786363235746385920
Gather id: 786286427768250368
Gather id: 786233965241827333
Gather id: 786051337297522688
Gather id: 786036967502913536
Gather id: 785927819176054784
Gather id: 785872687017132033
Gather id: 785639753186217984
Gather id: 785533386513321988
Gather id: 785515384317313025
Gather id: 785264754247995392
Gather id: 785170936622350336
Gather id: 784826020293709826
Gather id: 784517518371221505
Gather id: 784431430411685888
Gather id: 784183165795655680
Gather id: 784057939640352768
```

```
Gather id: 783839966405230592
Gather id: 783821107061198850
Gather id: 783695101801398276
Gather id: 783466772167098368
Gather id: 783391753726550016
Gather id: 783347506784731136
Gather id: 783334639985389568
Gather id: 783085703974514689
Gather id: 782969140009107456
Gather id: 782747134529531904
Gather id: 782722598790725632
Gather id: 782598640137187329
Gather id: 782305867769217024
Gather id: 782021823840026624
Gather id: 781955203444699136
Gather id: 781661882474196992
Gather id: 781655249211752448
Gather id: 781524693396357120
Gather id: 781308096455073793
Gather id: 781251288990355457
Gather id: 781163403222056960
Gather id: 780931614150983680
Gather id: 780858289093574656
Gather id: 780800785462489090
Gather id: 780601303617732608
Gather id: 780543529827336192
Gather id: 780496263422808064
Gather id: 780476555013349377
Gather id: 780459368902959104
Gather id: 780192070812196864
Gather id: 780092040432480260
Gather id: 780074436359819264
Gather id: 779834332596887552
Gather id: 779377524342161408
Gather id: 779124354206535695
Gather id: 779123168116150273
Error - id: 779123168116150273[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 779056095788752897
Gather id: 778990705243029504
Gather id: 778774459159379968
Gather id: 778764940568104960
Gather id: 778748913645780993
Gather id: 778650543019483137
```

```
Gather id: 778624900596654080
Gather id: 778408200802557953
Gather id: 778396591732486144
Gather id: 778383385161035776
Gather id: 778286810187399168
Gather id: 778039087836069888
Gather id: 778027034220126208
Gather id: 777953400541634568
Gather id: 777885040357281792
Gather id: 777684233540206592
Gather id: 777641927919427584
Gather id: 777621514455814149
Gather id: 777189768882946048
Gather id: 776819012571455488
Gather id: 776813020089548800
Gather id: 776477788987613185
Gather id: 776249906839351296
Gather id: 776218204058357768
Gather id: 776201521193218049
Gather id: 776113305656188928
Gather id: 776088319444877312
Gather id: 775898661951791106
Gather id: 775842724423557120
Gather id: 775733305207554048
Gather id: 775729183532220416
Gather id: 775364825476165632
Gather id: 775350846108426240
Gather id: 775096608509886464
Error - id: 775096608509886464[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 775085132600442880
Gather id: 774757898236878852
Gather id: 774639387460112384
Gather id: 774314403806253056
Gather id: 773985732834758656
Gather id: 773922284943896577
Gather id: 773704687002451968
Gather id: 773670353721753600
Gather id: 773547596996571136
Gather id: 773336787167145985
Gather id: 773308824254029826
Gather id: 773247561583001600
Gather id: 773191612633579521
Gather id: 772877495989305348
```

Gather id: 772826264096874500
Gather id: 772615324260794368
Gather id: 772581559778025472
Gather id: 772193107915964416
Gather id: 772152991789019136
Gather id: 772117678702071809
Gather id: 772114945936949249
Gather id: 772102971039580160
Gather id: 771908950375665664
Gather id: 771770456517009408
Gather id: 771500966810099713
Gather id: 771380798096281600
Gather id: 771171053431250945
Gather id: 771136648247640064
Gather id: 771102124360998913
Gather id: 771014301343748096
Gather id: 771004394259247104
Error - id: 771004394259247104[{'code': 179, 'message': 'Sorry, you are not authorized
to see this status.'}]
Gather id: 770787852854652928
Gather id: 770772759874076672
Gather id: 770743923962707968
Error - id: 770743923962707968[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 770655142660169732
Gather id: 770414278348247044
Gather id: 770293558247038976
Gather id: 770093767776997377
Gather id: 770069151037685760
Gather id: 769940425801170949
Gather id: 769695466921623552
Gather id: 769335591808995329
Gather id: 769212283578875904
Gather id: 768970937022709760
Gather id: 768909767477751808
Gather id: 768855141948723200
Gather id: 768609597686943744
Gather id: 768596291618299904
Gather id: 768554158521745409
Gather id: 768473857036525572
Gather id: 768193404517830656
Gather id: 767884188863397888
Gather id: 767754930266464257
Gather id: 767500508068192258

```
Gather id: 767191397493538821
Gather id: 767122157629476866
Gather id: 766864461642756096
Gather id: 766793450729734144
Gather id: 766714921925144576
Gather id: 766693177336135680
Gather id: 766423258543644672
Gather id: 766313316352462849
Gather id: 766078092750233600
Gather id: 766069199026450432
Gather id: 766008592277377025
Gather id: 765719909049503744
Gather id: 765669560888528897
Gather id: 765395769549590528
Gather id: 765371061932261376
Gather id: 765222098633691136
Gather id: 764857477905154048
Gather id: 764259802650378240
Gather id: 763956972077010945
Gather id: 763837565564780549
Gather id: 763183847194451968
Gather id: 763167063695355904
Gather id: 763103485927849985
Gather id: 762699858130116608
Gather id: 762471784394268675
Gather id: 762464539388485633
Gather id: 762316489655476224
Gather id: 762035686371364864
Gather id: 761976711479193600
Gather id: 761750502866649088
Gather id: 761745352076779520
Gather id: 761672994376806400
Gather id: 761599872357261312
Gather id: 761371037149827077
Gather id: 761334018830917632
Gather id: 761292947749015552
Gather id: 761227390836215808
Gather id: 761004547850530816
Gather id: 760893934457552897
Gather id: 760656994973933572
Gather id: 760641137271070720
Gather id: 760539183865880579
Gather id: 760521673607086080
Gather id: 760290219849637889
```

```
Gather id: 760252756032651264
Gather id: 760190180481531904
Gather id: 760153949710192640
Gather id: 759943073749200896
Gather id: 759923798737051648
Gather id: 759846353224826880
Gather id: 759793422261743616
Gather id: 759566828574212096
Error - id: 759566828574212096[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 759557299618865152
Gather id: 759447681597108224
Gather id: 759446261539934208
Gather id: 759197388317847553
Gather id: 759159934323924993
Gather id: 759099523532779520
Gather id: 759047813560868866
Gather id: 758854675097526272
Gather id: 758828659922702336
Gather id: 758740312047005698


Rate limit reached. Sleeping for: 755


Gather id: 758474966123810816
Gather id: 758467244762497024
Gather id: 758405701903519748
Gather id: 758355060040593408
Gather id: 758099635764359168
Gather id: 758041019896193024
Gather id: 757741869644341248
Gather id: 757729163776290825
Gather id: 757725642876129280
Gather id: 757611664640446465
Gather id: 757597904299253760
Gather id: 757596066325864448
Gather id: 757400162377592832
Gather id: 757393109802180609
Gather id: 757354760399941633
Gather id: 756998049151549440
Gather id: 756939218950160384
Gather id: 756651752796094464
Gather id: 756526248105566208
Gather id: 756303284449767430
Gather id: 756288534030475264
Gather id: 756275833623502848
```

```
Gather id: 755955933503782912
Gather id: 755206590534418437
Gather id: 755110668769038337
Gather id: 754874841593970688
Gather id: 754856583969079297
Gather id: 754747087846248448
Gather id: 754482103782404096
Gather id: 754449512966619136
Gather id: 754120377874386944
Gather id: 754011816964026368
Error - id: 754011816964026368[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 753655901052166144
Gather id: 753420520834629632
Gather id: 753398408988139520
Gather id: 753375668877008896
Gather id: 753298634498793472
Gather id: 753294487569522689
Gather id: 753039830821511168
Gather id: 753026973505581056
Gather id: 752932432744185856
Gather id: 752917284578922496
Gather id: 752701944171524096
Gather id: 752682090207055872
Gather id: 752660715232722944
Gather id: 752568224206688256
Gather id: 752519690950500352
Gather id: 752334515931054080
Gather id: 752309394570878976
Gather id: 752173152931807232
Gather id: 751950017322246144
Gather id: 751937170840121344
Gather id: 751830394383790080
Gather id: 751793661361422336
Gather id: 751598357617971201
Gather id: 751583847268179968
Gather id: 751538714308972544
Gather id: 751456908746354688
Gather id: 751251247299190784
Gather id: 751205363882532864
Gather id: 751132876104687617
Gather id: 750868782890057730
Gather id: 750719632563142656
Gather id: 750506206503038976
```

```
Gather id: 750429297815552001
Gather id: 750383411068534784
Gather id: 750381685133418496
Gather id: 750147208377409536
Gather id: 750132105863102464
Gather id: 750117059602808832
Gather id: 750101899009982464
Gather id: 750086836815486976
Gather id: 750071704093859840
Gather id: 750056684286914561
Gather id: 750041628174217216
Gather id: 750026558547456000
Gather id: 750011400160841729
Gather id: 749996283729883136
Gather id: 749981277374128128
Gather id: 749774190421639168
Gather id: 749417653287129088
Gather id: 749403093750648834
Gather id: 749395845976588288
Gather id: 749317047558017024
Gather id: 749075273010798592
Gather id: 749064354620928000
Gather id: 749036806121881602
Gather id: 748977405889503236
Gather id: 748932637671223296
Gather id: 748705597323898880
Gather id: 748699167502000129
Gather id: 748692773788876800
Gather id: 748575535303884801
Gather id: 748568946752774144
Gather id: 748346686624440324
Gather id: 748337862848962560
Gather id: 748324050481647620
Gather id: 748307329658011649
Gather id: 748220828303695873
Gather id: 747963614829678593
Gather id: 747933425676525569
Gather id: 747885874273214464
Gather id: 747844099428986880
Gather id: 747816857231626240
Gather id: 747651430853525504
Gather id: 747648653817413632
Gather id: 747600769478692864
Gather id: 747594051852075008
```

```
Gather id: 747512671126323200
Gather id: 747461612269887489
Gather id: 747439450712596480
Gather id: 747242308580548608
Gather id: 747219827526344708
Gather id: 747204161125646336
Gather id: 747103485104099331
Gather id: 746906459439529985
Gather id: 746872823977771008
Gather id: 746818907684614144
Gather id: 746790600704425984
Gather id: 746757706116112384
Gather id: 746726898085036033
Gather id: 746542875601690625
Gather id: 746521445350707200
Gather id: 746507379341139972
Gather id: 746369468511756288
Gather id: 746131877086527488
Gather id: 746056683365994496
Gather id: 745789745784041472
Gather id: 745712589599014916
Gather id: 745433870967832576
Gather id: 745422732645535745
Gather id: 745314880350101504
Gather id: 745074613265149952
Gather id: 745057283344719872
Gather id: 744995568523612160
Gather id: 744971049620602880
Gather id: 744709971296780288
Gather id: 744334592493166593
Gather id: 744234799360020481
Gather id: 744223424764059648
Gather id: 743980027717509120
Gather id: 743895849529389061
Gather id: 743835915802583040
Gather id: 743609206067040256
Gather id: 743595368194129920
Gather id: 743545585370791937
Gather id: 743510151680958465
Gather id: 743253157753532416
Gather id: 743222593470234624
Gather id: 743210557239623680
Gather id: 742534281772302336
Gather id: 742528092657332225
```

```
Gather id: 742465774154047488
Gather id: 742423170473463808
Gather id: 742385895052087300
Gather id: 742161199639494656
Gather id: 742150209887731712
Gather id: 741793263812808706
Gather id: 741743634094141440
Gather id: 741438259667034112
Gather id: 741303864243200000
Gather id: 741099773336379392
Gather id: 741067306818797568
Gather id: 740995100998766593
Gather id: 740711788199743490
Gather id: 740699697422163968
Gather id: 740676976021798912
Gather id: 740373189193256964
Gather id: 740365076218183684
Gather id: 740359016048689152
Gather id: 740214038584557568
Gather id: 739979191639244800
Gather id: 739932936087216128
Gather id: 739844404073074688
Gather id: 739623569819336705
Gather id: 739606147276148736
Gather id: 739544079319588864
Gather id: 739485634323156992
Gather id: 739238157791694849
Gather id: 738891149612572673
Gather id: 738885046782832640
Gather id: 738883359779196928
Gather id: 738537504001953792
Gather id: 738402415918125056
Gather id: 738184450748633089
Gather id: 738166403467907072
Gather id: 738156290900254721
Gather id: 737826014890496000
Gather id: 737800304142471168
Gather id: 737678689543020544
Gather id: 737445876994609152
Gather id: 737322739594330112
Gather id: 737310737551491075
Gather id: 736736130620620800
Gather id: 736392552031657984
Gather id: 736365877722001409
```

```
Gather id: 736225175608430592
Gather id: 736010884653420544
Gather id: 735991953473572864
Gather id: 735648611367784448
Gather id: 735635087207878657
Gather id: 735274964362878976
Gather id: 735256018284875776
Gather id: 735137028879360001
Gather id: 734912297295085568
Gather id: 734787690684657664
Gather id: 734776360183431168
Gather id: 734559631394082816
Gather id: 733828123016450049
Gather id: 733822306246479872
Gather id: 733482008106668032
Gather id: 733460102733135873
Gather id: 733109485275860992
Gather id: 732732193018155009
Gather id: 732726085725589504
Gather id: 732585889486888962
Gather id: 732375214819057664
Gather id: 732005617171337216
Gather id: 731285275100512256
Gather id: 731156023742988288
Gather id: 730924654643314689
Gather id: 730573383004487680
Gather id: 730427201120833536
Gather id: 730211855403241472
Gather id: 730196704625098752
Gather id: 729854734790754305
Gather id: 729838605770891264
Gather id: 729823566028484608
Gather id: 729463711119904772
Gather id: 729113531270991872
Gather id: 728986383096946689
Gather id: 728760639972315136
Gather id: 728751179681943552
Gather id: 728653952833728512
Gather id: 728409960103686147
Gather id: 728387165835677696
Gather id: 728046963732717569
Gather id: 728035342121635841
Gather id: 728015554473250816
Gather id: 727685679342333952
```

Gather id: 727644517743104000
Gather id: 727524757080539137
Gather id: 727314416056803329
Gather id: 727286334147182592
Gather id: 727175381690781696
Gather id: 727155742655025152
Gather id: 726935089318363137
Gather id: 726887082820554753
Gather id: 726828223124897792
Gather id: 726224900189511680
Gather id: 725842289046749185
Gather id: 725786712245440512
Gather id: 725729321944506368
Gather id: 725458796924002305
Gather id: 724983749226668032
Gather id: 724771698126512129
Gather id: 724405726123311104
Gather id: 724049859469295616
Gather id: 724046343203856385
Gather id: 724004602748780546
Gather id: 723912936180330496
Gather id: 723688335806480385
Gather id: 723673163800948736
Gather id: 723179728551723008
Gather id: 722974582966214656
Gather id: 722613351520608256
Gather id: 721503162398597120
Gather id: 721001180231503872
Gather id: 720785406564900865
Gather id: 720775346191278080
Gather id: 720415127506415616
Gather id: 720389942216527872
Gather id: 720340705894408192
Gather id: 720059472081784833
Gather id: 720043174954147842
Gather id: 719991154352222208
Gather id: 719704490224398336
Gather id: 719551379208073216
Gather id: 719367763014393856
Gather id: 719339463458033665
Gather id: 719332531645071360
Gather id: 718971898235854848
Gather id: 718939241951195136
Gather id: 718631497683582976

```
Gather id: 718613305783398402
Gather id: 718540630683709445
Gather id: 718460005985447936
Gather id: 718454725339934721
Gather id: 718246886998687744
Gather id: 718234618122661888
Gather id: 717841801130979328
Gather id: 717790033953034240
Gather id: 717537687239008257
Gather id: 717428917016076293
Gather id: 717421804990701568
Gather id: 717047459982213120
Gather id: 717009362452090881
Gather id: 716802964044845056
Gather id: 716791146589110272
Gather id: 716730379797970944
Gather id: 716447146686459905
Gather id: 716439118184652801
Gather id: 716285507865542656
Gather id: 716080869887381504
Gather id: 715928423106027520
Gather id: 715758151270801409
Gather id: 715733265223708672
Gather id: 715704790270025728
Gather id: 715696743237730304
Gather id: 715680795826982913
Gather id: 715360349751484417
Gather id: 715342466308784130
Gather id: 715220193576927233
Gather id: 715200624753819648
Gather id: 715009755312439296
Gather id: 714982300363173890
Gather id: 714962719905021952
Gather id: 714957620017307648
Gather id: 714631576617938945
Gather id: 714606013974974464
Gather id: 714485234495041536
Gather id: 714258258790387713
Gather id: 714251586676113411
Gather id: 714214115368108032
Gather id: 714141408463036416
Gather id: 713919462244790272
Gather id: 713909862279876608
Gather id: 713900603437621249
```

```
Gather id: 713761197720473600
Gather id: 713411074226274305
Gather id: 713177543487135744
Gather id: 713175907180089344
Gather id: 712809025985978368
Gather id: 712717840512598017
Gather id: 712668654853337088
Gather id: 712438159032893441
Gather id: 712309440758808576
Gather id: 712097430750289920
Gather id: 712092745624633345
Gather id: 712085617388212225
Gather id: 712065007010385924
Gather id: 711998809858043904
Gather id: 711968124745228288
Gather id: 711743778164514816
Gather id: 711732680602345472
Gather id: 711694788429553666
Gather id: 711652651650457602
Gather id: 711363825979756544
Gather id: 711306686208872448
Gather id: 711008018775851008
Gather id: 710997087345876993
Gather id: 710844581445812225
Gather id: 710833117892898816
Gather id: 710658690886586372
Gather id: 710609963652087808
Gather id: 710588934686908417
Gather id: 710296729921429505
Gather id: 710283270106132480
Gather id: 710272297844797440
Gather id: 710269109699739648
Gather id: 710153181850935296
Gather id: 710140971284037632
Gather id: 710117014656950272
Gather id: 709918798883774466
Gather id: 709901256215666688
Gather id: 709852847387627521
Gather id: 709566166965075968
Gather id: 709556954897764353
Gather id: 709519240576036864
Gather id: 709449600415961088
Gather id: 709409458133323776
Gather id: 709225125749587968
```

```
Gather id: 709207347839836162
Gather id: 709198395643068416
Gather id: 709179584944730112
Gather id: 709158332880297985
Gather id: 709042156699303936
Gather id: 708853462201716736
Gather id: 708845821941387268
Gather id: 708834316713893888
Gather id: 708810915978854401
Gather id: 708738143638450176
Gather id: 708711088997666817
Gather id: 708479650088034305
Gather id: 708469915515297792
Gather id: 708400866336894977
Gather id: 708356463048204288
Gather id: 708349470027751425
Gather id: 708149363256774660
Gather id: 708130923141795840
Gather id: 708119489313951744
Gather id: 708109389455101952
Gather id: 708026248782585858
Gather id: 707995814724026368
Gather id: 707983188426153984
Gather id: 707969809498152960
Gather id: 707776935007539200
Gather id: 707741517457260545
Gather id: 707738799544082433
Gather id: 707693576495472641
Gather id: 707629649552134146
Gather id: 707610948723478529
Gather id: 707420581654872064
Gather id: 707411934438625280
Gather id: 707387676719185920
Gather id: 707377100785885184
Gather id: 707315916783140866
Gather id: 707297311098011648
Gather id: 707059547140169728
Gather id: 707038192327901184
Gather id: 707021089608753152
Gather id: 707014260413456384
Gather id: 706904523814649856
Gather id: 706901761596989440
Gather id: 706681918348251136
Gather id: 706644897839910912
```

```
Gather id: 706593038911545345
Gather id: 706538006853918722
Gather id: 706516534877929472
Gather id: 706346369204748288
Gather id: 706310011488698368
Gather id: 706291001778950144
Gather id: 706265994973601792
Gather id: 706169069255446529
Gather id: 706166467411222528
Gather id: 706153300320784384
Gather id: 705975130514706432
Gather id: 705970349788291072
Gather id: 705898680587526145
Gather id: 705786532653883392
Gather id: 705591895322394625
Gather id: 705475953783398401
Gather id: 705442520700944385
Gather id: 705428427625635840
Gather id: 705239209544720384
Gather id: 705223444686888960
Gather id: 705102439679201280
Gather id: 705066031337840642
Gather id: 704871453724954624
Gather id: 704859558691414016
Gather id: 704847917308362754
Gather id: 704819833553219584
Gather id: 704761120771465216
Gather id: 704499785726889984
Gather id: 704491224099647488
Gather id: 704480331685040129
Gather id: 704364645503647744
Gather id: 704347321748819968
Gather id: 704134088924532736
Gather id: 704113298707505153
Gather id: 704054845121142784
Gather id: 703774238772166656
Gather id: 703769065844768768
Gather id: 703631701117943808
Gather id: 703611486317502464
Gather id: 703425003149250560
Gather id: 703407252292673536
Gather id: 703382836347330562
Gather id: 703356393781329922
Gather id: 703268521220972544
```

```
Gather id: 703079050210877440
Gather id: 703041949650034688
Gather id: 702932127499816960
Gather id: 702899151802126337
Gather id: 702684942141153280
Gather id: 702671118226825216
Gather id: 702598099714314240
Gather id: 702539513671897089
Gather id: 702332542343577600
Gather id: 702321140488925184
Gather id: 702276748847800320
Gather id: 702217446468493312
Gather id: 701981390485725185
Gather id: 701952816642965504
Gather id: 701889187134500865
Gather id: 701805642395348998
Gather id: 701601587219795968
Gather id: 701570477911896070
Gather id: 701545186879471618
Gather id: 701214700881756160
Gather id: 700890391244103680
Gather id: 700864154249383937
Gather id: 700847567345688576
Gather id: 700796979434098688
Gather id: 700747788515020802
Gather id: 700518061187723268
Gather id: 700505138482569216
Gather id: 700462010979500032
Gather id: 700167517596164096
Gather id: 700151421916807169
Gather id: 700143752053182464
Gather id: 700062718104104960
Gather id: 700029284593901568
Gather id: 700002074055016451
Gather id: 699801817392291840
Gather id: 699788877217865730
Gather id: 699779630832685056
Gather id: 699775878809702401
Gather id: 699691744225525762
Gather id: 699446877801091073
Gather id: 699434518667751424
Gather id: 699423671849451520
Gather id: 699413908797464576
Gather id: 699370870310113280
```

```
Gather id: 699323444782047232
Gather id: 699088579889332224
Gather id: 699079609774645248
Gather id: 699072405256409088
Gather id: 699060279947165696
Gather id: 699036661657767936
Gather id: 698989035503689728
Gather id: 698953797952008193
Gather id: 698907974262222848
Gather id: 698710712454139905
Gather id: 698703483621523456
Gather id: 698635131305795584
Gather id: 698549713696649216
Gather id: 698355670425473025
Gather id: 698342080612007937
Gather id: 698262614669991936
Gather id: 698195409219559425
Gather id: 698178924120031232
Gather id: 697995514407682048
Gather id: 697990423684476929
Gather id: 697943111201378304
Gather id: 697881462549430272
Gather id: 697630435728322560
Gather id: 697616773278015490
Gather id: 697596423848730625
Gather id: 697575480820686848
Gather id: 697516214579523584
Gather id: 697482927769255936
Gather id: 697463031882764288
Gather id: 697270446429966336
Gather id: 697259378236399616
Gather id: 697255105972801536
Gather id: 697242256848379904
Gather id: 696900204696625153
Gather id: 696894894812565505
Gather id: 696886256886657024
Gather id: 696877980375769088
Gather id: 696754882863349760
Gather id: 696744641916489729
Gather id: 696713835009417216
Gather id: 696518437233913856
Gather id: 696490539101908992
Gather id: 696488710901260288
Gather id: 696405997980676096
```

```
Gather id: 696100768806522880
Gather id: 695816827381944320
Gather id: 695794761660297217
Gather id: 695767669421768709
Gather id: 695629776980148225
Gather id: 695446424020918272
Gather id: 695409464418041856
Gather id: 695314793360662529
Gather id: 695095422348574720
Gather id: 695074328191332352
Gather id: 695064344191721472
Gather id: 695051054296211456
Gather id: 694925794720792577
Gather id: 694905863685980160
Gather id: 694669722378485760
Gather id: 694356675654983680
Gather id: 694352839993344000
Gather id: 694342028726001664
Gather id: 694329668942569472
Gather id: 694206574471057408
Gather id: 694183373896572928
Gather id: 694001791655137281
Gather id: 693993230313091072
Gather id: 693942351086120961
Gather id: 693647888581312512
Gather id: 693644216740769793
Gather id: 693642232151285760
Gather id: 693629975228977152
Gather id: 693622659251335168
Gather id: 693590843962331137
Gather id: 693582294167244802
Gather id: 693486665285931008
Gather id: 693280720173801472
Gather id: 693267061318012928
Gather id: 693262851218264065
Gather id: 693231807727280129
Gather id: 693155686491000832
Gather id: 693109034023534592
Gather id: 693095443459342336
Gather id: 692919143163629568
Gather id: 692905862751522816
Gather id: 692901601640583168
Gather id: 692894228850999298
Gather id: 692828166163931137
```

Gather id: 692752401762250755
Gather id: 692568918515392513
Gather id: 692535307825213440
Gather id: 692530551048294401
Gather id: 692423280028966913
Gather id: 692417313023332352
Gather id: 692187005137076224
Gather id: 692158366030913536
Gather id: 692142790915014657
Gather id: 692041934689402880
Gather id: 692017291282812928
Gather id: 691820333922455552
Gather id: 691793053716221953
Gather id: 691756958957883396
Gather id: 691675652215414786
Gather id: 691483041324204033
Gather id: 691459709405118465
Gather id: 691444869282295808
Gather id: 691416866452082688
Gather id: 691321916024623104
Gather id: 691096613310316544
Gather id: 691090071332753408
Gather id: 690989312272396288
Gather id: 690959652130045952
Gather id: 690938899477221376
Gather id: 690932576555528194
Gather id: 690735892932222976
Gather id: 690728923253055490
Gather id: 690690673629138944
Gather id: 690649993829576704
Gather id: 690607260360429569
Gather id: 690597161306841088
Gather id: 690400367696297985
Gather id: 690374419777196032
Gather id: 690360449368465409
Gather id: 690348396616552449
Gather id: 690248561355657216
Gather id: 690021994562220032
Gather id: 690015576308211712
Gather id: 690005060500217858
Gather id: 689999384604450816
Gather id: 689993469801164801
Gather id: 689977555553848577
Gather id: 689905486972461056

```
Gather id: 689877686181715968
Gather id: 689835978131935233
Gather id: 689661964914655233
Gather id: 689659372465688576
Gather id: 689623661272240129
Gather id: 689599056876867584
Gather id: 689557536375177216
Gather id: 689517482558820352
Gather id: 689289219123089408
Gather id: 689283819090870273
Gather id: 689280876073582592
Gather id: 689275259254616065
Gather id: 689255633275777024
Gather id: 689154315265683456
Gather id: 689143371370250240
Gather id: 688916208532455424
Gather id: 688908934925697024
Gather id: 688898160958271489
Gather id: 688894073864884227
Gather id: 688828561667567616
Gather id: 688804835492233216
Gather id: 688789766343622656
Gather id: 688547210804498433
Gather id: 688519176466644993
Gather id: 688385280030670848
Gather id: 688211956440801280
Gather id: 688179443353796608
Gather id: 688116655151435777
Gather id: 688064179421470721
Gather id: 687841446767013888
Gather id: 687826841265172480
Gather id: 687818504314159109
Gather id: 687807801670897665
Gather id: 687732144991551489
Gather id: 687704180304273409
Gather id: 687664829264453632
Gather id: 687494652870668288
Gather id: 687480748861947905
Gather id: 687476254459715584
Gather id: 687460506001633280
Gather id: 687399393394311168
Gather id: 687317306314240000
Gather id: 687312378585812992
Gather id: 687127927494963200
```

```
Gather id: 687124485711986689
Gather id: 687109925361856513
Gather id: 687102708889812993
Gather id: 687096057537363968
Gather id: 686947101016735744
Gather id: 686760001961103360
Gather id: 686749460672679938
Gather id: 686730991906516992
Gather id: 686683045143953408
Gather id: 686618349602762752
Gather id: 686606069955735556
Gather id: 686394059078897668
Gather id: 686386521809772549
Gather id: 686377065986265092
Gather id: 686358356425093120
Gather id: 686286779679375361
Gather id: 686050296934563840
Gather id: 686035780142297088
Gather id: 686034024800862208
Gather id: 686007916130873345
Gather id: 686003207160610816
Gather id: 685973236358713344
Gather id: 685943807276412928
Gather id: 685906723014619143
Gather id: 685681090388975616
Gather id: 685667379192414208
Gather id: 685663452032069632
Gather id: 685641971164143616
Gather id: 685547936038666240
Gather id: 685532292383666176
Gather id: 685325112850124800
Gather id: 685321586178670592
Gather id: 685315239903100929
Gather id: 685307451701334016
Gather id: 685268753634967552
Gather id: 685198997565345792
Gather id: 685169283572338688
Gather id: 684969860808454144
Gather id: 684959798585110529
Gather id: 684940049151070208
Gather id: 684926975086034944
Gather id: 684914660081053696
Gather id: 684902183876321280
Gather id: 684880619965411328
```

Gather id: 684830982659280897
Gather id: 684800227459624960
Gather id: 684594889858887680
Gather id: 684588130326986752
Gather id: 684567543613382656
Gather id: 684538444857667585
Gather id: 684481074559381504
Gather id: 684460069371654144
Gather id: 684241637099323392
Gather id: 684225744407494656
Gather id: 684222868335505415
Gather id: 684200372118904832
Gather id: 684195085588783105
Gather id: 684188786104872960
Gather id: 684177701129875456
Gather id: 684147889187209216
Gather id: 684122891630342144
Gather id: 684097758874210310
Gather id: 683857920510050305
Gather id: 683852578183077888
Gather id: 683849932751646720
Gather id: 683834909291606017
Gather id: 683828599284170753
Gather id: 683773439333797890
Gather id: 683742671509258241
Gather id: 683515932363329536
Gather id: 683498322573824003
Gather id: 683481228088049664
Gather id: 683462770029932544
Gather id: 683449695444799489
Gather id: 683391852557561860
Gather id: 683357973142474752
Gather id: 683142553609318400
Gather id: 683111407806746624
Gather id: 683098815881154561
Gather id: 683078886620553216
Gather id: 683030066213818368
Gather id: 682962037429899265
Gather id: 682808988178739200
Gather id: 682788441537560576
Gather id: 682750546109968385
Gather id: 682697186228989953
Gather id: 682662431982772225
Gather id: 682638830361513985

```
Gather id: 682429480204398592
Gather id: 682406705142087680
Gather id: 682393905736888321
Gather id: 682389078323662849
Gather id: 682303737705140231
Gather id: 682259524040966145
Gather id: 682242692827447297
Gather id: 682088079302213632
Gather id: 682059653698686977
Gather id: 682047327939461121
Gather id: 682032003584274432
Gather id: 682003177596559360
Gather id: 681981167097122816
Gather id: 681891461017812993
Gather id: 681694085539872773
Gather id: 681679526984871937
Gather id: 681654059175129088
Gather id: 681610798867845120
Gather id: 681579835668455424
Gather id: 681523177663676416
Gather id: 681340665377193984
Gather id: 681339448655802368
Gather id: 681320187870711809
Gather id: 681302363064414209
Gather id: 681297372102656000
Gather id: 681281657291280384
Gather id: 681261549936340994
Gather id: 681242418453299201
Gather id: 681231109724700672
Gather id: 681193455364796417
Gather id: 680970795137544192
Gather id: 680959110691590145
Gather id: 680940246314430465
Gather id: 680934982542561280
Gather id: 680913438424612864
Gather id: 680889648562991104
Gather id: 680836378243002368
Gather id: 680805554198020098
Gather id: 680801747103793152
Gather id: 680798457301471234
Gather id: 680609293079592961
Gather id: 680583894916304897
Gather id: 680497766108381184
Gather id: 680494726643068929
```

```
Gather id: 680473011644985345
Gather id: 680440374763077632
Gather id: 680221482581123072
Gather id: 680206703334408192
Gather id: 680191257256136705
Gather id: 680176173301628928
Gather id: 680161097740095489
Gather id: 680145970311643136
Gather id: 680130881361686529
Gather id: 680115823365742593
Gather id: 680100725817409536
Gather id: 680085611152338944
Gather id: 680070545539371008
Gather id: 680055455951884288
Error - id: 680055455951884288[{'code': 144, 'message': 'No status found with that
ID.'}]
Gather id: 679877062409191424
Gather id: 679872969355714560
Gather id: 679862121895714818
Gather id: 679854723806179328
Gather id: 679844490799091713
Gather id: 679828447187857408
Gather id: 679777920601223168
Gather id: 679736210798047232
Gather id: 679729593985699840
Gather id: 679722016581222400
Gather id: 679530280114372609
Gather id: 679527802031484928
Gather id: 679511351870550016
Gather id: 679503373272485890
Gather id: 679475951516934144
Gather id: 679462823135686656
Gather id: 679405845277462528
Gather id: 679158373988876288
Gather id: 679148763231985668
Gather id: 679132435750195208
Gather id: 679111216690831360
Gather id: 679062614270468097
Gather id: 679047485189439488
Gather id: 679001094530465792
Gather id: 678991772295516161
Gather id: 678969228704284672
Gather id: 678800283649069056
Gather id: 678798276842360832
```

```
Gather id: 678774928607469569
Gather id: 678767140346941444
Gather id: 678764513869611008
Gather id: 678755239630127104
Gather id: 678740035362037760
Gather id: 678708137298427904
Gather id: 678675843183484930
Gather id: 678643457146150913
Gather id: 678446151570427904
Gather id: 678424312106393600
Gather id: 678410210315247616
Gather id: 678399652199309312
Gather id: 678396796259975168
Gather id: 678389028614488064
Gather id: 678380236862578688
Gather id: 678341075375947776
Gather id: 678334497360859136
Gather id: 678278586130948096
Gather id: 678255464182861824
Gather id: 678023323247357953
Gather id: 678021115718029313
Gather id: 677961670166224897
Gather id: 677918531514703872
Gather id: 677895101218201600
Gather id: 677716515794329600
Gather id: 677700003327029250
Gather id: 677698403548192770
Gather id: 677687604918272002
Gather id: 677673981332312066
Gather id: 677662372920729601
Gather id: 677644091929329666
Gather id: 677573743309385728
Gather id: 677565715327688705
Gather id: 677557565589463040
Gather id: 677547928504967168
Gather id: 677530072887205888
Gather id: 677335745548390400
Gather id: 677334615166730240
Gather id: 677331501395156992
Gather id: 677328882937298944
Gather id: 677314812125323265
Gather id: 677301033169788928
Gather id: 677269281705472000
Gather id: 677228873407442944
```

```
Gather id: 677187300187611136
Gather id: 676975532580409345

Rate limit reached. Sleeping for: 757

Gather id: 676957860086095872
Gather id: 676949632774234114
Gather id: 676948236477857792
Gather id: 676946864479084545
Gather id: 676942428000112642
Gather id: 676936541936185344
Gather id: 676916996760600576
Gather id: 676897532954456065
Gather id: 676864501615042560
Gather id: 676821958043033607
Gather id: 676819651066732545
Gather id: 676811746707918848
Gather id: 676776431406465024
Gather id: 676617503762681856
Gather id: 676613908052996102
Gather id: 676606785097199616
Gather id: 676603393314578432
Gather id: 676593408224403456
Gather id: 676590572941893632
Gather id: 676588346097852417
Gather id: 676582956622721024
Gather id: 676575501977128964
Gather id: 676533798876651520
Gather id: 676496375194980353
Gather id: 676470639084101634
Gather id: 676440007570247681
Gather id: 676430933382295552
Gather id: 676263575653122048
Gather id: 676237365392908289
Gather id: 676219687039057920
Gather id: 676215927814406144
Gather id: 676191832485810177
Gather id: 676146341966438401
Gather id: 676121918416756736
Gather id: 676101918813499392
Gather id: 676098748976615425
Gather id: 676089483918516224
Gather id: 675898130735476737
Gather id: 675891555769696257
Gather id: 675888385639251968
```

```
Gather id: 675878199931371520
Gather id: 675870721063669760
Gather id: 675853064436391936
Gather id: 675849018447167488
Gather id: 675845657354215424
Gather id: 675822767435051008
Gather id: 675820929667219457
Gather id: 675798442703122432
Gather id: 675781562965868544
Gather id: 675740360753160193
Gather id: 675710890956750848
Gather id: 675707330206547968
Gather id: 675706639471788032
Gather id: 675534494439489536
Gather id: 675531475945709568
Gather id: 675522403582218240
Gather id: 675517828909424640
Gather id: 675501075957489664
Gather id: 675497103322386432
Gather id: 675489971617296384
Gather id: 675483430902214656
Gather id: 675432746517426176
Gather id: 675372240448454658
Gather id: 675362609739206656
Gather id: 675354435921575936
Gather id: 675349384339542016
Gather id: 675334060156301312
Gather id: 675166823650848770
Gather id: 675153376133427200
Gather id: 675149409102012420
Gather id: 675147105808306176
Gather id: 675146535592706048
Gather id: 675145476954566656
Gather id: 675135153782571009
Gather id: 675113801096802304
Gather id: 675111688094527488
Gather id: 675109292475830276
Gather id: 675047298674663426
Gather id: 675015141583413248
Gather id: 675006312288268288
Gather id: 675003128568291329
Gather id: 674999807681908736
Gather id: 674805413498527744
Gather id: 674800520222154752
```

```
Gather id: 674793399141146624
Gather id: 674790488185167872
Gather id: 674788554665512960
Gather id: 674781762103414784
Gather id: 674774481756377088
Gather id: 674767892831932416
Gather id: 674764817387900928
Gather id: 674754018082705410
Gather id: 674752233200820224
Gather id: 674743008475090944
Gather id: 674742531037511680
Gather id: 674739953134403584
Gather id: 674737130913071104
Gather id: 674690135443775488
Gather id: 674670581682434048
Gather id: 674664755118911488
Gather id: 674646392044941312
Gather id: 674644256330530816
Gather id: 674638615994089473
Gather id: 674632714662858753
Gather id: 674606911342424069
Gather id: 674468880899788800
Gather id: 674447403907457024
Gather id: 674436901579923456
Gather id: 674422304705744896
Gather id: 674416750885273600
Gather id: 674410619106390016
Gather id: 674394782723014656
Gather id: 674372068062928900
Gather id: 674330906434379776
Gather id: 674318007229923329
Gather id: 674307341513269249
Gather id: 674291837063053312
Gather id: 674271431610523648
Gather id: 674269164442398721
Gather id: 674265582246694913
Gather id: 674262580978937856
Gather id: 674255168825880576
Gather id: 674082852460433408
Gather id: 674075285688614912
Gather id: 674063288070742018
Gather id: 674053186244734976
Gather id: 674051556661161984
Gather id: 674045139690631169
```

```
Gather id: 674042553264685056
Gather id: 674038233588723717
Gather id: 674036086168010753
Gather id: 674024893172875264
Gather id: 674019345211760640
Gather id: 674014384960745472
Gather id: 674008982932058114
Gather id: 673956914389192708
Gather id: 673919437611909120
Gather id: 673906403526995968
Gather id: 673887867907739649
Gather id: 673716320723169284
Gather id: 673715861853720576
Gather id: 673711475735838725
Gather id: 673709992831262724
Gather id: 673708611235921920
Gather id: 673707060090052608
Gather id: 673705679337693185
Gather id: 673700254269775872
Gather id: 673697980713705472
Gather id: 673689733134946305
Gather id: 673688752737402881
Gather id: 673686845050527744
Gather id: 673680198160809984
Gather id: 673662677122719744
Gather id: 673656262056419329
Gather id: 673636718965334016
Gather id: 673612854080196609
Gather id: 673583129559498752
Gather id: 673580926094458881
Gather id: 673576835670777856
Gather id: 673363615379013632
Gather id: 673359818736984064
Gather id: 673355879178194945
Gather id: 673352124999274496
Gather id: 673350198937153538
Gather id: 673345638550134785
Gather id: 673343217010679808
Gather id: 673342308415348736
Gather id: 673320132811366400
Gather id: 673317986296586240
Gather id: 673295268553605120
Gather id: 673270968295534593
Gather id: 673240798075449344
```

```
Gather id: 673213039743795200
Gather id: 673148804208660480
Gather id: 672997845381865473
Gather id: 672995267319328768
Gather id: 672988786805112832
Gather id: 672984142909456390
Gather id: 672980819271634944
Gather id: 672975131468300288
Gather id: 672970152493887488
Gather id: 672968025906282496
Gather id: 672964561327235073
Gather id: 672902681409806336
Gather id: 672898206762672129
Gather id: 672884426393653248
Gather id: 672877615439593473
Gather id: 672834301050937345
Gather id: 672828477930868736
Gather id: 672640509974827008
Gather id: 672622327801233409
Gather id: 672614745925664768
Gather id: 672609152938721280
Gather id: 672604026190569472
Gather id: 672594978741354496
Gather id: 672591762242805761
Gather id: 672591271085670400
Gather id: 672538107540070400
Gather id: 672523490734551040
Gather id: 672488522314567680
Gather id: 672482722825261057
Gather id: 672481316919734272
Gather id: 672475084225949696
Gather id: 672466075045466113
Gather id: 672272411274932228
Gather id: 672267570918129665
Gather id: 672264251789176834
Gather id: 672256522047614977
Gather id: 672254177670729728
Gather id: 672248013293752320
Gather id: 672245253877968896
Gather id: 672239279297454080
Gather id: 672231046314901505
Gather id: 672222792075620352
Gather id: 672205392827572224
Gather id: 672169685991993344
```

```
Gather id: 672160042234327040
Gather id: 672139350159835138
Gather id: 672125275208069120
Gather id: 672095186491711488
Gather id: 672082170312290304
Gather id: 672068090318987265
Gather id: 671896809300709376
Gather id: 671891728106971137
Gather id: 671882082306625538
Gather id: 671879137494245376
Gather id: 671874878652489728
Gather id: 671866342182637568
Gather id: 671855973984772097
Gather id: 671789708968640512
Gather id: 671768281401958400
Gather id: 671763349865160704
Gather id: 671744970634719232
Gather id: 671743150407421952
Gather id: 671735591348891648
Gather id: 671729906628341761
Gather id: 671561002136281088
Gather id: 671550332464455680
Gather id: 671547767500775424
Gather id: 671544874165002241
Gather id: 671542985629241344
Gather id: 671538301157904385
Gather id: 671536543010570240
Gather id: 671533943490011136
Gather id: 671528761649688577
Gather id: 671520732782923777
Gather id: 671518598289059840
Gather id: 671511350426865664
Gather id: 671504605491109889
Gather id: 671497587707535361
Gather id: 671488513339211776
Gather id: 671486386088865792
Gather id: 671485057807351808
Gather id: 671390180817915904
Gather id: 671362598324076544
Gather id: 671357843010908160
Gather id: 671355857343524864
Gather id: 671347597085433856
Gather id: 671186162933985280
Gather id: 671182547775299584
```

Gather id: 671166507850801152
Gather id: 671163268581498880
Gather id: 671159727754231808
Gather id: 671154572044468225
Gather id: 671151324042559489
Gather id: 671147085991960577
Gather id: 671141549288370177
Gather id: 671138694582165504
Gather id: 671134062904504320
Gather id: 671122204919246848
Gather id: 671115716440031232
Gather id: 671109016219725825
Gather id: 670995969505435648
Gather id: 670842764863651840
Gather id: 670840546554966016
Gather id: 670838202509447168
Gather id: 670833812859932673
Gather id: 670832455012716544
Gather id: 670826280409919488
Gather id: 670823764196741120
Gather id: 670822709593571328
Gather id: 670815497391357952
Gather id: 670811965569282048
Gather id: 670807719151067136
Gather id: 670804601705242624
Gather id: 670803562457407488
Gather id: 670797304698376195
Gather id: 670792680469889025
Gather id: 670789397210615808
Gather id: 670786190031921152
Gather id: 670783437142401025
Gather id: 670782429121134593
Gather id: 670780561024270336
Gather id: 670778058496974848
Gather id: 670764103623966721
Gather id: 670755717859713024
Gather id: 670733412878163972
Gather id: 670727704916926465
Gather id: 670717338665226240
Gather id: 670704688707301377
Gather id: 670691627984359425
Gather id: 670679630144274432
Gather id: 670676092097810432
Gather id: 670668383499735048

```
Gather id: 670474236058800128
Gather id: 670468609693655041
Gather id: 670465786746662913
Gather id: 670452855871037440
Gather id: 670449342516494336
Gather id: 670444955656130560
Gather id: 670442337873600512
Gather id: 670435821946826752
Gather id: 670434127938719744
Gather id: 670433248821026816
Gather id: 670428280563085312
Gather id: 670427002554466305
Gather id: 670421925039075328
Gather id: 670420569653809152
Gather id: 670417414769758208
Gather id: 670411370698022913
Gather id: 670408998013820928
Gather id: 670403879788544000
Gather id: 670385711116361728
Gather id: 670374371102445568
Gather id: 670361874861563904
Gather id: 670338931251150849
Gather id: 670319130621435904
Gather id: 670303360680108032
Gather id: 670290420111441920
Gather id: 670093938074779648
Gather id: 670086499208155136
Gather id: 670079681849372674
Gather id: 670073503555706880
Gather id: 670069087419133954
Gather id: 670061506722140161
Gather id: 670055038660800512
Gather id: 670046952931721218
Gather id: 670040295598354432
Gather id: 670037189829525505
Gather id: 670003130994700288
Gather id: 669993076832759809
Gather id: 669972011175813120
Gather id: 669970042633789440
Gather id: 669942763794931712
Gather id: 669926384437997569
Gather id: 669923323644657664
Gather id: 669753178989142016
Gather id: 669749430875258880
```

```
Gather id: 669684865554620416
Gather id: 669683899023405056
Gather id: 669682095984410625
Gather id: 669680153564442624
Gather id: 669661792646373376
Gather id: 669625907762618368
Gather id: 669603084620980224
Gather id: 669597912108789760
Gather id: 669583744538451968
Gather id: 669573570759163904
Gather id: 669571471778410496
Gather id: 669567591774625800
Gather id: 669564461267722241
Gather id: 669393256313184256
Gather id: 669375718304980992
Gather id: 669371483794317312
Gather id: 669367896104181761
Gather id: 669363888236994561
Gather id: 669359674819481600
Gather id: 669354382627049472
Gather id: 669353438988365824
Gather id: 669351434509529089
Gather id: 669328503091937280
Gather id: 669327207240699904
Gather id: 669324657376567296
Gather id: 669216679721873412
Gather id: 669214165781868544
Gather id: 669203728096960512
Gather id: 669037058363662336
Gather id: 669015743032369152
Gather id: 669006782128353280
Gather id: 669000397445533696
Gather id: 668994913074286592
Gather id: 668992363537309700
Gather id: 668989615043424256
Gather id: 668988183816871936
Gather id: 668986018524233728
Gather id: 668981893510119424
Gather id: 668979806671884288
Gather id: 668975677807423489
Gather id: 668967877119254528
Gather id: 668960084974809088
Gather id: 668955713004314625
Gather id: 668932921458302977
```

```
Gather id: 668902994700836864
Gather id: 668892474547511297
Gather id: 668872652652679168
Gather id: 668852170888998912
Gather id: 668826086256599040
Gather id: 668815180734689280
Gather id: 668779399630725120
Gather id: 668655139528511488
Gather id: 668645506898350081
Gather id: 668643542311546881
Gather id: 668641109086707712
Gather id: 668636665813057536
Gather id: 668633411083464705
Gather id: 668631377374486528
Gather id: 668627278264475648
Gather id: 668625577880875008
Gather id: 668623201287675904
Gather id: 668620235289837568
Gather id: 668614819948453888
Gather id: 668587383441514497
Gather id: 668567822092664832
Gather id: 668544745690562560
Gather id: 668542336805281792
Gather id: 668537837512433665
Gather id: 668528771708952576
Gather id: 668507509523615744
Gather id: 668496999348633600
Gather id: 668484198282485761
Gather id: 668480044826800133
Gather id: 668466899341221888
Gather id: 668297328638447616
Gather id: 668291999406125056
Gather id: 668286279830867968
Gather id: 668274247790391296
Gather id: 668268907921326080
Gather id: 668256321989451776
Gather id: 668248472370458624
Gather id: 668237644992782336
Gather id: 668226093875376128
Gather id: 668221241640230912
Gather id: 668204964695683073
Gather id: 668190681446379520
Gather id: 668171859951755264
Gather id: 668154635664932864
```

```
Gather id: 668142349051129856
Gather id: 668113020489474048
Gather id: 667937095915278337
Gather id: 667924896115245057
Gather id: 667915453470232577
Gather id: 667911425562669056
Gather id: 667902449697558528
Gather id: 667886921285246976
Gather id: 667885044254572545
Gather id: 667878741721415682
Gather id: 667873844930215936
Gather id: 667866724293877760
Gather id: 667861340749471744
Gather id: 667832474953625600
Gather id: 667806454573760512
Gather id: 667801013445750784
Gather id: 667793409583771648
Gather id: 667782464991965184
Gather id: 667773195014021121
Gather id: 667766675769573376
Gather id: 667728196545200128
Gather id: 667724302356258817
Gather id: 667550904950915073
Gather id: 667550882905632768
Gather id: 667549055577362432
Gather id: 667546741521195010
Gather id: 667544320556335104
Gather id: 667538891197542400
Gather id: 667534815156183040
Gather id: 667530908589760512
Gather id: 667524857454854144
Gather id: 667517642048163840
Gather id: 667509364010450944
Gather id: 667502640335572993
Gather id: 667495797102141441
Gather id: 667491009379606528
Gather id: 667470559035432960
Gather id: 667455448082227200
Gather id: 667453023279554560
Gather id: 667443425659232256
Gather id: 667437278097252352
Gather id: 667435689202614272
Gather id: 667405339315146752
Gather id: 667393430834667520
```

```
Gather id: 667369227918143488
Gather id: 667211855547486208
Gather id: 667200525029539841
Gather id: 667192066997374976
Gather id: 667188689915760640
Gather id: 667182792070062081
Gather id: 667177989038297088
Gather id: 667176164155375616
Gather id: 667174963120574464
Gather id: 667171260800061440
Gather id: 667165590075940865
Gather id: 667160273090932737
Gather id: 667152164079423490
Gather id: 667138269671505920
Gather id: 667119796878725120
Gather id: 667090893657276420
Gather id: 667073648344346624
Gather id: 667070482143944705
Gather id: 667065535570550784
Gather id: 667062181243039745
Gather id: 667044094246576128
Gather id: 667012601033924608
Gather id: 666996132027977728
Gather id: 666983947667116034
Gather id: 666837028449972224
Gather id: 666835007768551424
Gather id: 666826780179869698
Gather id: 666817836334096384
Gather id: 666804364988780544
Gather id: 666786068205871104
Gather id: 666781792255496192
Gather id: 666776908487630848
Gather id: 666739327293083650
Gather id: 666701168228331520
Gather id: 666691418707132416
Gather id: 666649482315059201
Gather id: 666644823164719104
Gather id: 666454714377183233
Gather id: 666447344410484738
Gather id: 666437273139982337
Gather id: 666435652385423360
Gather id: 666430724426358785
Gather id: 666428276349472768
Gather id: 666421158376562688
```

```
Gather id: 666418789513326592
Gather id: 666411507551481857
Gather id: 666407126856765440
Gather id: 666396247373291520
Gather id: 666373753744588802
Gather id: 666362758909284353
Gather id: 666353288456101888
Gather id: 666345417576210432
Gather id: 666337882303524864
Gather id: 666293911632134144
Gather id: 666287406224695296
Gather id: 666273097616637952
Gather id: 666268910803644416
Gather id: 666104133288665088
Gather id: 666102155909144576
Gather id: 666099513787052032
Gather id: 666094000022159362
Gather id: 666082916733198337
Gather id: 666073100786774016
Gather id: 666071193221509120
Gather id: 666063827256086533
Gather id: 666058600524156928
Gather id: 666057090499244032
Gather id: 666055525042405380
Gather id: 666051853826850816
Gather id: 666050758794694657
Gather id: 666049248165822465
Gather id: 666044226329800704
Gather id: 666033412701032449
Gather id: 666029285002620928
Gather id: 666020888022790149
```

In [0]:

```python
api_data = []

#read the created file
with open("tweet_json.txt", "r") as f:
    for line in f:
        try:
            tweet = json.loads(line)
            #append a dictionary to the created list
            api_data.append({
                "tweet_id": tweet["id"],
                "retweet_count": tweet["retweet_count"],
```

```
                "favorite_count": tweet["favorite_count"],
                "retweeted": tweet["retweeted"],
                "display_text_range": tweet["display_text_range"]
            })

            #tweet["entities"]["media"][0]["media_url"]
        except:
            print("Error.")

df_api = pd.DataFrame(api_data, columns = ["tweet_id", "retweet_count",
"favorite_count", "retweeted", "display_text_range"])
df_api.head()
```

Out[0]:

	tweet_id	retweet_count	favorite_count	retweeted	display_text_range
0	892420643555336193	7723	36281	False	[0, 85]
1	892177421306343426	5712	31283	False	[0, 138]
2	891815181378084864	3786	23553	False	[0, 121]
3	891689557279858688	7878	39567	False	[0, 79]
4	891327558926688256	8498	37794	False	[0, 138]

Thus the JSON file table is created

In [0]:

```
df_twitter.head(1)
```

Out[0]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	retweeted_status_est
0	892420643555336193	NaN	NaN	2017-08-01 16:23:56 +0000	<a href="http://twitter.com/download/iphone" r...	This is Phineas. He's a mystical boy. Only eve...	NaN	NaN	NaN

In [0]:

```
df_predict.head(1)
```

Out[0]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3
0	666020888022790149	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg	1	Welsh_springer_spaniel	0.465074	True	collie	0.156665	True	Shetland sheepd

In [0]:

```
df_api.head(1)
```

Out[0]:

	tweet_id	retweet_count	favorite_count	retweeted	display_text_range
0	892420643555336193	7723	36281	False	[0, 85]

Data Assessing¶

Now we will look at all the dataset and see where we need to clean the data

df_twitter

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_twitter.isnull(), vmin=0, vmax = 1)
```
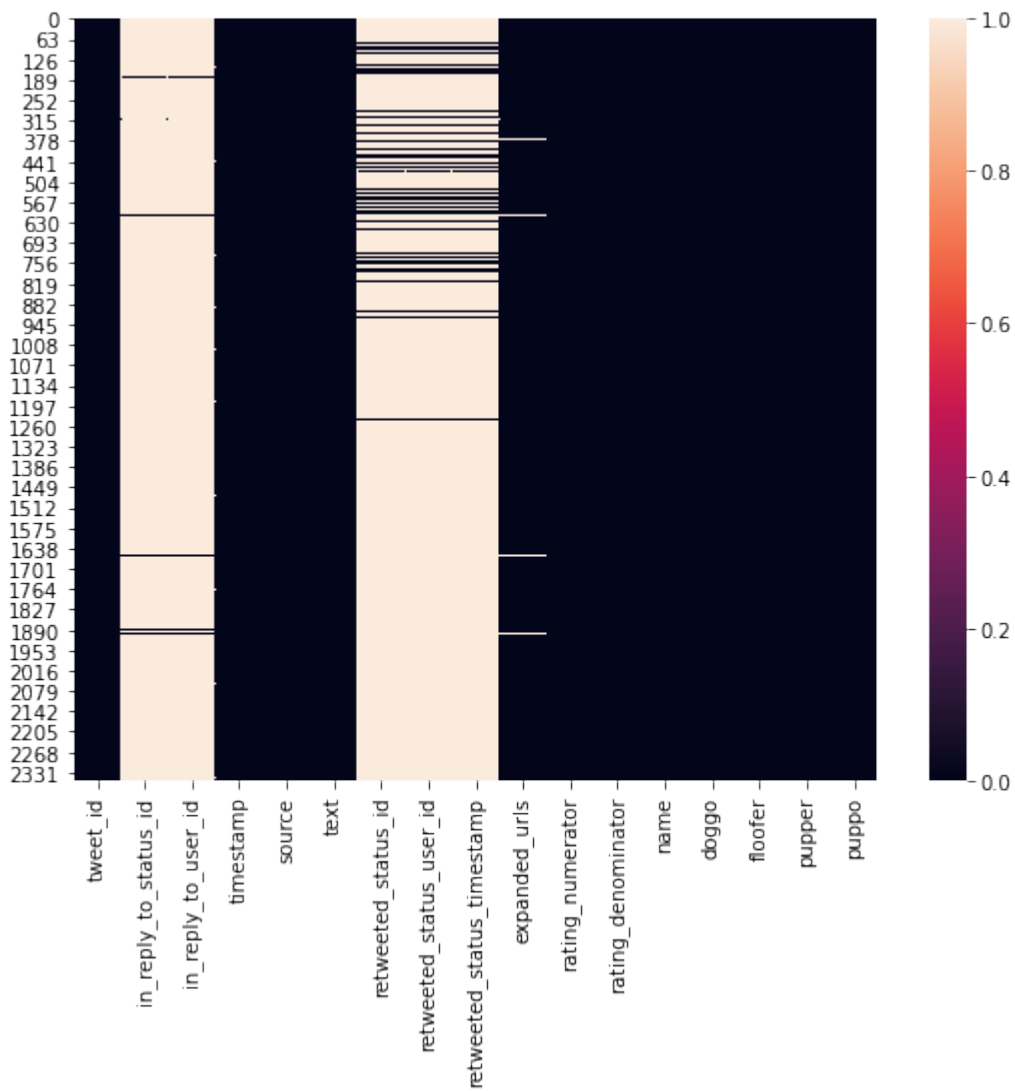
Here from the visual we can see that there a lot of missing datas in the reply and the retweet section

In [0]:

```
df_twitter.sample(5)
```

Out[0]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	re... s...
1366	702671118226825216	NaN	NaN	2016-02-25 01:47:04 +0000	<a href="http://t witter.com/do wnload/iphon e" r...	Meet Rambo & Kiwi. Rambo's the pup with th...	NaN	NaN	N...
2228	66825632198945177	NaN	NaN	2015-11-22	<a	This is	NaN	NaN	N...

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	re s
	6			02:34:57 +0000	href="http://t witter.com/do wnload/iphon e" r...	Jareld. Jareld rules these waters. Lad...			
657	79177493146595328 0	NaN	NaN	2016-10-27 22:53:48 +0000	<a href="http:// vine.co" rel="nofollow ">Vine -...	Vine will be deeply missed. This was by far my...	NaN	NaN	Na
2000	67248852231456768 0	NaN	NaN	2015-12-03 18:52:12 +0000	<a href="http://t witter.com/do wnload/iphon e" r...	This is Jackie. She was all ready to go out, b...	NaN	NaN	Na
555	80369222323786547 2	NaN	NaN	2016-11-29 20:08:52 +0000	<a href="http://t witter.com/do wnload/iphon e" r...	RT @dog_ra tes: I present to you... Dog Jesus. ...	6.914169e+ 17	4.196984e+ 09	2(0(+

here we can see that the doggo ,floofer , pupper ,puppo has a majority value of None .

In [0]:

df_twitter.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   tweet_id                2356 non-null   int64
 1   in_reply_to_status_id   78 non-null     float64
 2   in_reply_to_user_id     78 non-null     float64
 3   timestamp               2356 non-null   object
```

```
 4   source                      2356 non-null   object
 5   text                        2356 non-null   object
 6   retweeted_status_id          181 non-null   float64
 7   retweeted_status_user_id     181 non-null   float64
 8   retweeted_status_timestamp   181 non-null   object
 9   expanded_urls               2297 non-null   object
10   rating_numerator            2356 non-null   int64
11   rating_denominator          2356 non-null   int64
12   name                        2356 non-null   object
13   doggo                       2356 non-null   object
14   floofer                     2356 non-null   object
15   pupper                      2356 non-null   object
16   puppo                       2356 non-null   object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

Also we can see that tbere is incorrect data type

1. tweet_id is int but it should be string
2. timestamp is object but it should be data and time

In [0]:

```
df_twitter.name.value_counts()
```

Out[0]:

```
None          745
a              55
Charlie        12
Lucy           11
Oliver         11
           ...
Naphaniel       1
Emma            1
Kaia            1
Laela           1
Ralphé          1
Name: name, Length: 957, dtype: int64
```

Here we can see that there are many wrong name like "None", "a", "the" etc.

In [0]:

```
#to check the duplicated data
df_twitter[df_twitter.duplicated()]
```

Out[0]:

tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	retweeted_status_timestamp	expanded_urls	ratin me

we can see that there is no duplicated data

In [0]:

```
#how many unique tweet id
df_twitter.tweet_id.nunique()
```

Out[0]:

2356

This also proves that there are no duplicated values since the entire data set has unique tweet_ids

In [0]:

```
df_twitter_assess = df_twitter.copy()
```

In [0]:

```
#returns true if there is no dog classification in any of the columns
df_twitter_assess[["doggo","floofer","pupper","puppo"]].apply(lambda x: True if
    (x[0] == "None" and x[1] == "None" and x[2] == "None" and x[3] =="None")
    else False, axis = 1).value_counts()
```

Out[0]:

```
True     1976
False     380
dtype: int64
```

In [0]:

```
df_twitter_assess["doggo"].value_counts()
```

Out[0]:

```
None     2259
doggo      97
Name: doggo, dtype: int64
```

In [0]:

```
df_twitter_assess["floofer"].value_counts()
```

Out[0]:

```
None      2346
floofer     10
```

Name: floofer, dtype: int64

In [0]:

```
df_twitter_assess["pupper"].value_counts()
```

Out[0]:

```
None      2099
pupper     257
Name: pupper, dtype: int64
```

In [0]:

```
df_twitter_assess["puppo"].value_counts()
```

Out[0]:

```
None      2326
puppo       30
Name: puppo, dtype: int64
```

In [0]:

```
(380/2356) * 100
```

Out[0]:

```
16.129032258064516
```

So we can see that only about 16% of the data is not missing.

Now lets have a look at the Ratings

In [0]:

```
df_twitter.rating_numerator.value_counts()
```

Out[0]:

```
12      558
11      464
10      461
13      351
9       158
8       102
7        55
14       54
5        37
6        32
3        19
4        17
```

```
1        9
2        9
420      2
0        2
15       2
75       2
80       1
20       1
24       1
26       1
44       1
50       1
60       1
165      1
84       1
88       1
144      1
182      1
143      1
666      1
960      1
1776     1
17       1
27       1
45       1
99       1
121      1
204      1
Name: rating_numerator, dtype: int64
```

Here we can see that the highest Rating is 1776 and the lowest is 0. So lets see them in details

In [0]:

```
print(df_twitter.query("rating_numerator == '1776'").text)
```

```
979     This is Atticus. He's quite simply America af....
Name: text, dtype: object
```

In [0]:

```
print(df_twitter.query("rating_numerator == '1'").text)
```

```
605     RT @dog_rates: Not familiar with this breed. N...
1446    After reading the comments I may have overesti...
1869    What kind of person sends in a picture without...
1940    The millennials have spoken and we've decided ...
```

```
2038    After 22 minutes of careful deliberation this ...
2091    Flamboyant pup here. Probably poisonous. Won't...
2261    Never seen dog like this. Breathes heavy. Tilt...
2335    This is an Albanian 3 1/2 legged  Episcopalian...
2338    Not familiar with this breed. No tail (weird)....
Name: text, dtype: object
```

Here we can see that 18669 sends a photo without a dog and 2335 is being misinterpreted

In [0]:

```
print(df_twitter.query("rating_numerator == '0'").text)
```

```
315     When you're so blinded by your systematic plag...
1016    PUPDATE: can't see any. Even if I could, I cou...
Name: text, dtype: object
```

We can see here the dogs photos are not clear

In [0]:

```
#now lets check the denominator values
df_twitter.rating_denominator.value_counts()
```

Out[0]:

```
10      2333
11         3
50         3
80         2
20         2
2          1
16         1
40         1
70         1
15         1
90         1
110        1
120        1
130        1
150        1
170        1
7          1
0          1
Name: rating_denominator, dtype: int64
```

We can see that the highest denominator is 170 and the lowest is zero

In [0]:

```python
print(df_twitter.query("rating_denominator == '170'").text)
```

```
1120    Say hello to this unbelievably well behaved sq...
Name: text, dtype: object
```

In [0]:

```python
print(df_twitter.query("rating_denominator == '0'").text)
```

```
313    @jonnysun @Lin_Manuel ok jomny I know you're e...
Name: text, dtype: object
```

In [0]:

```python
print(df_twitter.query("rating_denominator == '7'").text)
```

```
516    Meet Sam. She smiles 24/7 &amp; secretly aspir...
Name: text, dtype: object
```

In [0]:

```python
#this is to extract all the actual ratings and split the denom and num into seperate
columns
pattern = "(\d+(\.\d+)?\/\d+(\.\d+)?)"
df_twitter_assess["rating"] = df_twitter_assess.text.str.extract(pattern, expand =
True)[0]
df_twitter_assess[['num', 'denom']] = df_twitter_assess['rating'].str.split('/', n=1,
expand=True)
```

In [0]:

```python
df_twitter_assess.rating_numerator = df_twitter_assess.rating_numerator.astype("str")
df_twitter_assess.rating_denominator =
df_twitter_assess.rating_denominator.astype("str")
```

In [0]:

```python
df_twitter_assess["check_num"] = df_twitter_assess[["rating_numerator",
"num"]].apply(lambda x: False if (x[0] != x[1]) else True, axis = 1)
```

In [0]:

```python
df_twitter_assess.check_num.value_counts()
```

Out[0]:

```
True     2349
False       7
Name: check_num, dtype: int64
```

In [0]:

```
df_twitter_assess.query("check_num == False")[["rating_numerator", "num","check_num"]]
```

Out[0]:

	rating_numerator	num	check_num
45	5	13.5	False
340	75	9.75	False
387	7	007	False
695	75	9.75	False
763	27	11.27	False
1689	5	9.5	False
1712	26	11.26	False

These are the differences we found by extracting the first occurrence of the pattern. These ratings got transformed to integers and are therefore wrong.

In [0]:

```
df_twitter_assess[["rating_numerator", "num", "check_num"]].sample(15)
```

Out[0]:

	rating_numerator	num	check_num
583	11	11	True
952	11	11	True
459	13	13	True
1321	12	12	True
12	13	13	True
1976	8	8	True
1886	13	13	True
2315	8	8	True
351	12	12	True
623	13	13	True
962	11	11	True
1840	10	10	True
169	12	12	True
1363	6	6	True
1110	10	10	True

In [0]:

```
#doing the same for denominator
```

```
df_twitter_assess["check_denom"] = df_twitter_assess[["rating_denominator",
"denom"]].apply(lambda x: False if (x[0] != x[1]) else True, axis = 1)
```

In [0]:

```
df_twitter_assess.check_denom.value_counts()
```

Out[0]:

```
True      2355
False        1
Name: check_denom, dtype: int64
```

In [0]:

```
df_twitter_assess.query("check_denom == False")[["rating_denominator",
"denom","check_denom"]]
```

Out[0]:

	rating_denominator	denom	check_denom
313	0	00	False

Its not a issue. we can keep the data same

In [0]:

```
df_twitter_assess[["rating_denominator", "denom", "check_denom"]].sample(5)
```

Out[0]:

	rating_denominator	denom	check_denom
1677	10	10	True
987	10	10	True
1751	10	10	True
1794	10	10	True
1434	10	10	True

Now lets asssess the data to look how many times there are multiple ratings

In [0]:

```
df_twitter_assess["count"] = df_twitter_assess.text.str.count(pattern)
```

In [0]:

```
df_twitter_assess["count"].value_counts()
```

Out[0]:

```
1    2323
2      32
3       1
```

Name: count, dtype: int64

In [0]:

```
#show the full text
pd.set_option('display.max_columns', 500)
pd.set_option('display.max_colwidth', 4000)

df_twitter_assess[["text", "count"]].query("count != 1")
```

Out[0]:

	text	count
55	@roushfenway These are good dogs but 17/10 is an emotional impulse rating. More like 13/10s	2
313	@jonnysun @Lin_Manuel ok jomny I know you're excited but 960/00 isn't a valid rating, 13/10 is tho	2
561	RT @dog_rates: "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup not flossing https://t.co/SuXcI9B7pQ	2
766	"Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup not flossing https://t.co/SuXcI9B7pQ	2
784	RT @dog_rates: After so many requests, this is Bretagne. She was the last surviving 9/11 search dog, and our second ever 14/10. RIP https:/...	2
860	RT @dog_rates: Meet Eve. She's a raging alcoholic 8/10 (would b 11/10 but pupper alcoholism is a tragic issue that I can't condone) https:/...	2
1007	This is Bookstore and Seaweed. Bookstore is tired and Seaweed is an asshole. 10/10 and 7/10 respectively https://t.co/eUGjGjjFVJ	2
1068	After so many requests, this is Bretagne. She was the last surviving 9/11 search dog, and our second ever 14/10. RIP https://t.co/XAVDNDaVgQ	2
1165	Happy 4/20 from the squad! 13/10 for all https://t.co/eV1diwds8a	2
1202	This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amazed af. 11/10 https://t.co/Kky1DPG4iq	2
1222	Meet Travis and Flurp. Travis is pretty chill but Flurp can't lie down properly. 10/10 & 8/10\nget it together Flurp https://t.co/Akzl5ynMmE	2
1359	This is Socks. That water pup w the super legs just splashed him. Socks did not appreciate that. 9/10 and 2/10 https://t.co/8rc5I22bBf	2
1459	This may be the greatest video I've ever been sent. 4/10 for Charles the puppy, 13/10 overall. (Vid by @stevenxx_) https://t.co/uaJmNgXR2P	2
1465	Meet Oliviér. He takes killer selfies. Has a dog of his own. It leaps at random & can't bark for shit. 10/10 & 5/10 https://t.co/6NgsQJuSBJ	2
1508	When bae says they can't go out but you see them with someone else that same night. 5/10 & 10/10 for heartbroken pup https://t.co/aenk0KpoWM	2

	text	count
1525	This is Eriq. His friend just reminded him of last year's super bowl. Not cool friend\n10/10 for Eriq\n6/10 for friend https://t.co/PlEXTofdpf	2
1538	Meet Fynn & Taco. Fynn is an all-powerful leaf lord and Taco is in the wrong place at the wrong time. 11/10 & 10/10 https://t.co/MuqHPvtL8c	2
1662	This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was just spotted by the helicopter 10/10 https://t.co/7EsP8LmSp5	2
1795	Meet Tassy & Bee. Tassy is pretty chill, but Bee is convinced the Ruffles are haunted. 10/10 & 11/10 respectively https://t.co/fgORpmTN9C	2
1832	These two pups just met and have instantly bonded. Spectacular scene. Mesmerizing af. 10/10 and 7/10 for blue dog https://t.co/gwryaJO4tC	2
1897	Meet Rufio. He is unaware of the pink legless pupper wrapped around him. Might want to get that checked 10/10 & 4/10 https://t.co/KNfLnYPmYh	2
1901	Two gorgeous dogs here. Little waddling dog is a rebel. Refuses to look at camera. Must be a preteen. 5/10 & 8/10 https://t.co/YPfw7oahbD	2
1970	Meet Eve. She's a raging alcoholic 8/10 (would b 11/10 but pupper alcoholism is a tragic issue that I can't condone) https://t.co/U36HYQIijg	2
2010	10/10 for dog. 7/10 for cat. 12/10 for human. Much skill. Would pet all https://t.co/uhx5gfpx5k	3
2064	Meet Holly. She's trying to teach small human-like pup about blocks but he's not paying attention smh. 11/10 & 8/10 https://t.co/RcksaUrGNu	2
2113	Meet Hank and Sully. Hank is very proud of the pumpkin they found and Sully doesn't give a shit. 11/10 and 8/10 https://t.co/cwoP1ftbrj	2
2177	Here we have Pancho and Peaches. Pancho is a Condoleezza Gryffindor, and Peaches is just an asshole. 10/10 & 7/10 https://t.co/Lh1BsJrWPp	2
2216	This is Spark. He's nervous. Other dog hasn't moved in a while. Won't come when called. Doesn't fetch well 8/10&1/10 https://t.co/stEodX9Aba	2
2263	This is Kial. Kial is either wearing a cape, which would be rad, or flashing us, which would be rude. 10/10 or 4/10 https://t.co/8zcwIoiuqR	2
2272	Two dogs in this one. Both are rare Jujitsu Pythagoreans. One slightly whiter than other. Long legs. 7/10 and 8/10 https://t.co/ITxxcc4v9y	2
2298	After much debate this dog is being upgraded to 10/10. I repeat 10/10	2
2306	These are Peruvian Feldspars. Their names are Cupit and Prencer. Both resemble Rand Paul. Sick outfits 10/10 & 10/10 https://t.co/ZnEMHBsAs1	2
2335	This is an Albanian 3 1/2 legged Episcopalian. Loves well-polished hardwood flooring. Penis on the collar. 9/10 https://t.co/d9NcXFKwLv	2

We can see that:

this data contains retweets (as mentioned before)
sometimes there are multiple dogs/cats or else in one picture

some of these ratings are not clear

df_Predict

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_predict.isnull(), vmin = 0, vmax = 1)
```
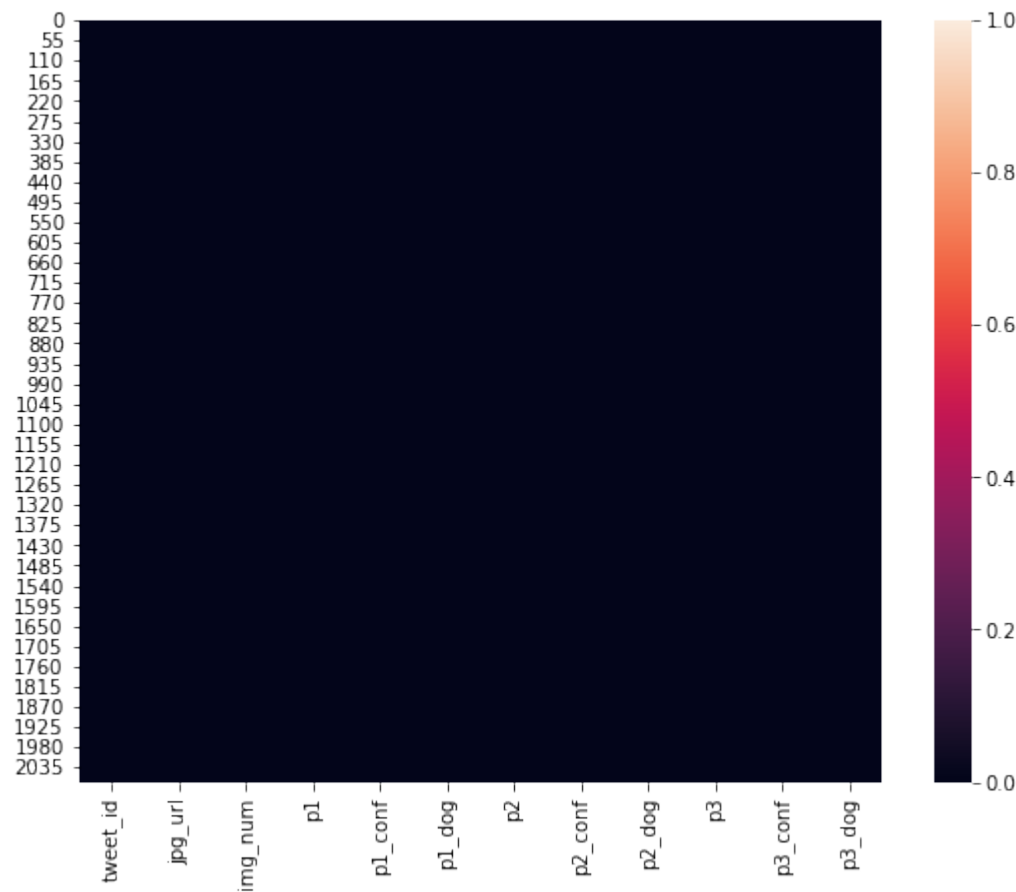

It looks perfect with no missing data

In [0]:

```
df_predict.sample(10)
```

Out[0]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_con
451	674743008475090944	https://pbs.twimg.com/media/CV0rL7RWEAAbhqm.jpg	1	Bernese_mountain_dog	0.583054	True	Shetland_sheepdog	0.06599
320	671789708968640512	https://pbs.twimg.com/tweet_video_thumb/CVKtH-4WIAAmiQ5.png	1	dalmatian	0.114259	True	teddy	0.06227
1793	83126262738074828	https://pbs.twimg.com/	1	cocker_spani	0.263323	True	Brittany_s	0.20055

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_con
	9	media/C4k88lGVMAEKNzb.jpg		el			paniel	
567	678341075375947776	https://pbs.twimg.com/media/CWnznDTU4AAa-6P.jpg	1	golden_retriever	0.853284	True	cocker_spaniel	0.02623
1749	823699002998870016	https://pbs.twimg.com/media/C25d3nkXEAAFBUN.jpg	1	cairn	0.203999	True	snorkel	0.17189
935	703407252292673536	https://pbs.twimg.com/media/CcMBJODUsAI5-A9.jpg	1	doormat	0.201058	False	turnstile	0.08858
292	671186162933985280	https://pbs.twimg.com/media/CVCIQX7UkAEzqh_.jpg	1	Chihuahua	0.319106	True	whippet	0.16913
690	684195085588783105	https://pbs.twimg.com/media/CX6_y6OU0AAl3v2.jpg	1	Chihuahua	0.379365	True	toy_terrier	0.12180
604	679854723806179328	https://pbs.twimg.com/media/CW9UQ7oWkAAErmU.jpg	1	llama	0.887963	False	ram	0.09502
36	666447344410484738	https://pbs.twimg.com/media/CT-yU5QWwAEjLX5.jpg	1	curly-coated_retriever	0.322084	True	giant_schnauzer	0.28795

We can see that:

1.the predicitions are sometimes lowercase, sometimes uppercase

1. there is an underscore instead of a whitespace between the words
2. there are rows with no prediciton of a dog (neither in 1, 2 nor 3)

In [0]:

```
df_predict.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   tweet_id   2075 non-null   int64
 1   jpg_url    2075 non-null   object
 2   img_num    2075 non-null   int64
 3   p1         2075 non-null   object
```

```
 4    p1_conf    2075 non-null    float64
 5    p1_dog     2075 non-null    bool
 6    p2         2075 non-null    object
 7    p2_conf    2075 non-null    float64
 8    p2_dog     2075 non-null    bool
 9    p3         2075 non-null    object
10    p3_conf    2075 non-null    float64
11    p3_dog     2075 non-null    bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

Here again the tweet id should be string

In [0]:

```
#to check the duplicated values
df_predict.jpg_url.value_counts().head(10)
```

Out[0]:

```
https://pbs.twimg.com/media/CU3mITUWIAAfyQS.jpg
2
https://pbs.twimg.com/media/Cbs3DOAXIAAp3Bd.jpg
2
https://pbs.twimg.com/media/CvoBPWRWgAA4het.jpg
2
https://pbs.twimg.com/media/CtKHLuCWYAA2TTs.jpg
2
https://pbs.twimg.com/ext_tw_video_thumb/817423809049493505/pu/img/5OFW0yueFu9oTUiQ.jpg
2
https://pbs.twimg.com/media/CVMOlMiWwAA4Yxl.jpg
2
https://pbs.twimg.com/ext_tw_video_thumb/815965888126062592/pu/img/JleSw4wRhgKDWQj5.jpg
2
https://pbs.twimg.com/media/Cveg1-NXgAASaaT.jpg
2
https://pbs.twimg.com/media/CdHwZd0VIAA4792.jpg
2
https://pbs.twimg.com/media/CtzKC7zXEAALfSo.jpg
2
Name: jpg_url, dtype: int64
```

In [0]:

```
df_predict[df_predict.tweet_id.duplicated()]
```

Out[0]:

tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog

In [0]:

```
df_predict[df_predict.jpg_url == "https://pbs.twimg.com/media/CzG425nWgAAnP7P.jpg"]
```

Out[0]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog
1638	806629075125202948	https://pbs.twimg.com/media/CzG425nWgAAnP7P.jpg	2	Arabian_camel	0.366248	False	house_finch	0.209852	False	c sp
1944	861769973181624320	https://pbs.twimg.com/media/CzG425nWgAAnP7P.jpg	2	Arabian_camel	0.366248	False	house_finch	0.209852	False	c sp

In [0]:

```
df_twitter[df_twitter.tweet_id ==861769973181624320]
```

Out[0]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	
155	861769973181624320	NaN	NaN	2017-05-09 02:29:07 +0000	Twitter for iPhone	RT @dog_rates: "Good afternoon class today we're going to learn what makes a good boy so good" 13/10 https://t.co/f1h2Fsalv9	8.066291e+17	4.196984e+09	2 2 +

In [0]:

```
df_predict.query("p1_dog == False and p2_dog == False and p3_dog == False")
```

Out[0]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p...
6	666051853826850816	https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg	1	box_turtle	0.933012	False	mud_turtle	0.045885	Fa
17	666104133288665088	https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg	1	hen	0.965932	False	cock	0.033919	Fa
18	666268910803644416	https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg	1	desktop_computer	0.086502	False	desk	0.085547	Fa
21	666293911632134144	https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg	1	three-toed_sloth	0.914671	False	otter	0.015250	Fa
25	666362758909284353	https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg	1	guinea_pig	0.996496	False	skunk	0.002402	Fa
...	...	...	...	...	...	...	...	...	...
2021	880935762899988482	https://pbs.twimg.com/media/DDm2Z5aXUAEDS2u.jpg	1	street_sign	0.251801	False	umbrella	0.115123	Fa
2022	881268444196462592	https://pbs.twimg.com/media/DDrk-f9WAAI-WQv.jpg	1	tusker	0.473303	False	Indian_elephant	0.245646	Fa
2046	886680336477933568	https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg	1	convertible	0.738995	False	sports_car	0.139952	Fa
2052	887517139158093824	https://pbs.twimg.com/ext_tw_video_thumb/887517108413886465/pu/img/WanJKwssZj4VJvL9.jpg	1	limousine	0.130432	False	tow_truck	0.029175	Fa
2074	892420643555336193	https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg	1	orange	0.097049	False	bagel	0.085851	Fa

324 rows × 12 columns

In [0]:

df_predict.query("p1_dog == False and (p2_dog == True or p3_dog == True)")

Out[0]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf
8	666057090499244032	https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg	1	shopping_cart	0.962465	False	shopping_basket	0.01459
22	666337882303524864	https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg	1	ox	0.416669	False	Newfoundland	0.27840
33	666430724426358785	https://pbs.twimg.com/media/CT-jNYqW4AAPi2M.jpg	1	llama	0.505184	False	Irish_terrier	0.10410
43	666776908487630848	https://pbs.twimg.com/media/CUDeDoWUYAAD-EM.jpg	1	seat_belt	0.375057	False	miniature_pinscher	0.16717
52	666996132027977728	https://pbs.twimg.com/media/CUGlb6iUwAITEbW.jpg	1	hay	0.507637	False	Rottweiler	0.06249
...	...	...	...	...	...	...	...	...
1984	872122724285648897	https://pbs.twimg.com/media/DBpm-5UXcAUeCru.jpg	1	basketball	0.808396	False	pug	0.06673
1992	873697596434513921	https://pbs.twimg.com/media/DA7iHL5U0AA1OQo.jpg	1	laptop	0.153718	False	French_bulldog	0.09998
2013	879376492567855104	https://pbs.twimg.com/media/DDQsQGFV0AAw6u9.jpg	1	tricycle	0.663601	False	Labrador_retriever	0.03349
2026	882045870035918850	https://pbs.twimg.com/media/DD2oCl2WAAEl_4a.jpg	1	web_site	0.949591	False	dhole	0.01732
2071	891689557279858688	https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg	1	paper_towel	0.170278	False	Labrador_retriever	0.16808

219 rows × 12 columns

After looking at some pictures it is clear that some dogs pictures are not properly visible

df_api

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_api.isnull(), vmin = 0, vmax = 1)
```
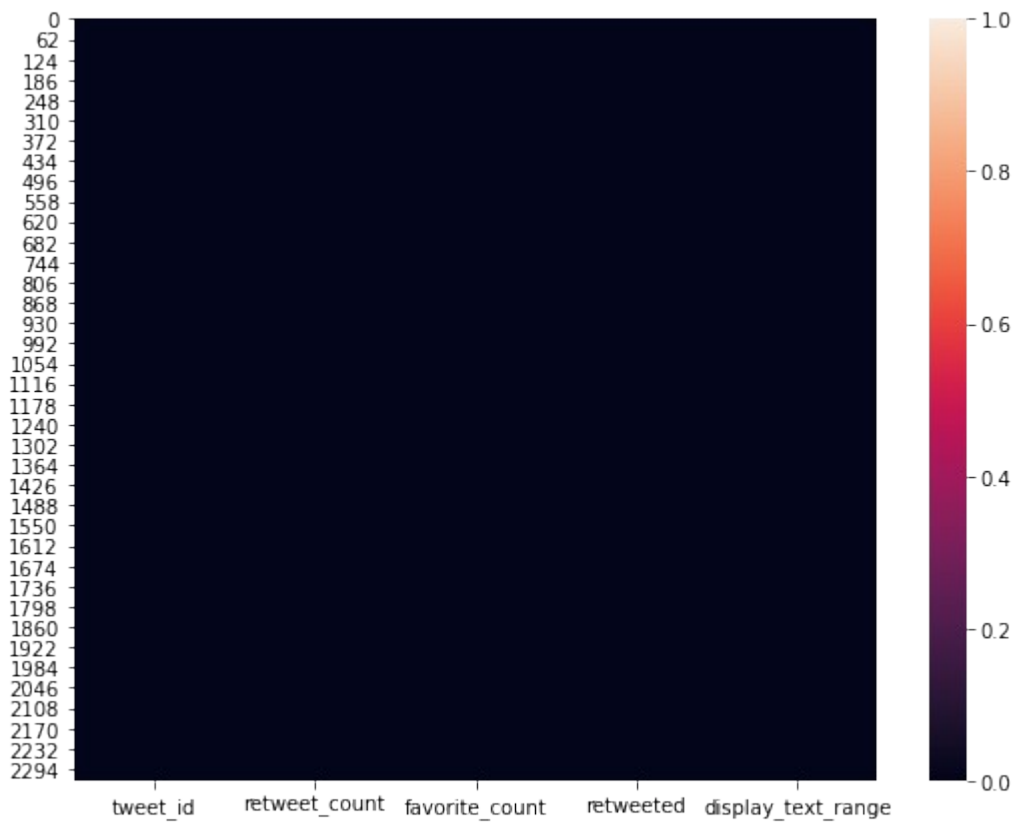
we can see that there is no missing values

In [0]:

df_api.sample(10)

Out[0]:

	tweet_id	retweet_count	favorite_count	retweeted	display_text_range
496	811647686436880384	760	5724	False	[0, 136]
1597	684926975086034944	460	3490	False	[0, 140]
169	857746408056729600	10171	33147	False	[0, 137]
611	794355576146903043	10575	0	False	[0, 118]
2223	667866724293877760	983	2910	False	[0, 140]
146	862722525377298433	3343	16405	False	[0, 135]
1997	672068090318987265	498	1250	False	[0, 138]
767	77463938746011238	1756	6832	False	[0, 76]

	tweet_id	retweet_count	favorite_count	retweeted	display_text_range
	4				
524	806242860592926720	11924	0	False	[0, 137]
970	748568946752774144	670	2203	False	[0, 85]

In [0]:

```
df_api.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2331 entries, 0 to 2330
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   tweet_id            2331 non-null   int64
 1   retweet_count       2331 non-null   int64
 2   favorite_count      2331 non-null   int64
 3   retweeted           2331 non-null   bool
 4   display_text_range  2331 non-null   object
dtypes: bool(1), int64(3), object(1)
memory usage: 75.2+ KB
```

This dataset looks okay. There is no missing values and only thing that needs to be changed is tweet id datatype to string

Assessing Summary¶

Quality¶

df_twitter table

1. the datatype of the id - columns is integer and should be str
2. the datatype of the timestamp - column is object and should be datetime
3. some of the dogs are not classified as one of "doggo", "floofer", "pupper" or "puppo" and contain all "None" instead
4. some of the dog names are not correct (None, an, by, a, ...) contains retweets
5. some of the ratings are not correctly extracted (mostly if there are >1 entries with the pattern "(\d+(.\d+)?\/\d+(.\d+)?)"
6. also transforming the ratings to integer created some mistakes (there are also floats)
7. the source column contains html code

df_predict table

```
1. the datatype of the id - columns is integer and should be str
```

2. contains retweets (duplicated rows in column jpg_url)
3. there are pictures in this table that are not dogs
4. the predictions are sometimes uppercase, sometimes lowercase
5. also there is a "_" instead of a whitespace in the predictions

df_api table

1. the datatype of the id - columns is integer and should be str

Tidiness¶

df_twitter table

1. the columns doggo, floofer,pupper and puppo are not easy to analyze and should be in one column

df_predict table

1. the prediction and confidence columns should be reduced to two columns - one for the prediction with the highest confidence (dog)

df_api table

    1. display_text_range contains 2 variables

all tables

1. All three tables share the column tweet_id and should be merged together.

Data Cleaning¶

Cleaning steps:

1. Merge the tables together
2. Drop the replies, retweets and the corresponding columns and also drop the tweets without an image or with images which don't display doggos
3. Clean the datatypes of the columns
4. Clean the wrong numerators - the floats on the one hand (replacement), the ones with multiple occurence of the pattern on the other (drop)
5. Extract the source from html code
6. Split the text range into two separate columns

7. Remove the "None" out of the doggo, floofer, pupper and puppo column and merge them into one column
8. Remove the wrong names of name column
9. Reduce the prediction columns into two - breed and conf
10. Clean the new breed column by replacing the "_" with a whitespace and make them all lowercase

Define:

merge the tables

code

In [0]:

```
#outer join to not loose rows at first
df_master = pd.merge(df_twitter, df_api, on = "tweet_id", how = "outer")
```

In [0]:

```
df_master = pd.merge(df_master, df_predict, on = "tweet_id", how = "outer")
```

In [0]:

```
df_master_clean = df_master.copy()
```

test

In [0]:

```
df_master_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 32 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   tweet_id                  2356 non-null   int64
 1   in_reply_to_status_id     78 non-null     float64
 2   in_reply_to_user_id       78 non-null     float64
 3   timestamp                 2356 non-null   object
 4   source                    2356 non-null   object
 5   text                      2356 non-null   object
 6   retweeted_status_id       181 non-null    float64
 7   retweeted_status_user_id  181 non-null    float64
 8   retweeted_status_timestamp 181 non-null   object
 9   expanded_urls             2297 non-null   object
 10  rating_numerator          2356 non-null   int64
 11  rating_denominator        2356 non-null   int64
 12  name                      2356 non-null   object
```

```
 13   doggo                            2356 non-null    object
 14   floofer                          2356 non-null    object
 15   pupper                           2356 non-null    object
 16   puppo                            2356 non-null    object
 17   retweet_count                    2331 non-null    float64
 18   favorite_count                   2331 non-null    float64
 19   retweeted                        2331 non-null    object
 20   display_text_range               2331 non-null    object
 21   jpg_url                          2075 non-null    object
 22   img_num                          2075 non-null    float64
 23   p1                               2075 non-null    object
 24   p1_conf                          2075 non-null    float64
 25   p1_dog                           2075 non-null    object
 26   p2                               2075 non-null    object
 27   p2_conf                          2075 non-null    float64
 28   p2_dog                           2075 non-null    object
 29   p3                               2075 non-null    object
 30   p3_conf                          2075 non-null    float64
 31   p3_dog                           2075 non-null    object
dtypes: float64(10), int64(3), object(19)
memory usage: 607.4+ KB
```
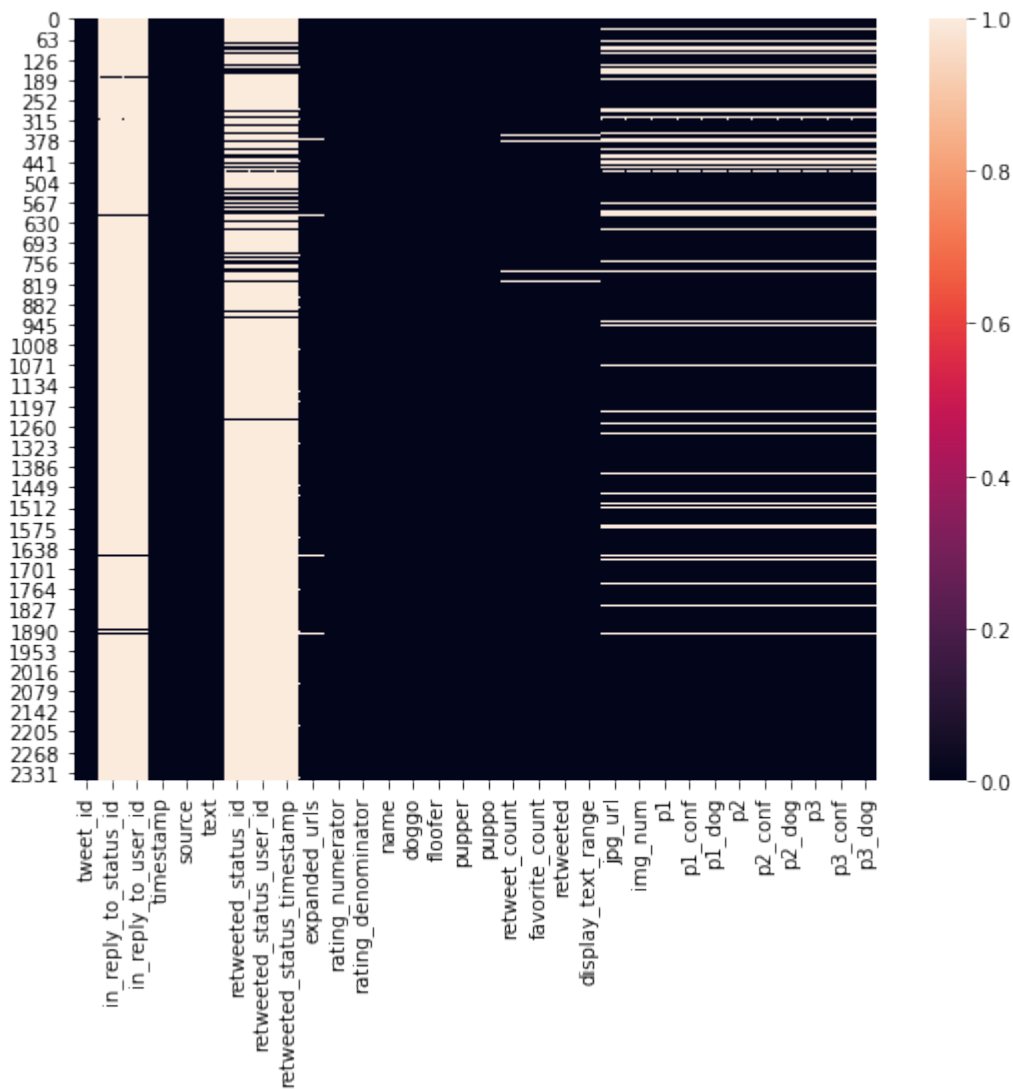
Define:

Drop the replies, retweets and the corresponding columns and also drop the tweets without an image or with images which don't display doggos

Code:

we can see that there are a lot of missing values...let's visualise it

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_master_clean.isnull(), vmin = 0, vmax = 1)
```

In [0]:

```
pd.set_option('display.max_colwidth', 50) #to set the column width
df_master_clean[df_master_clean["retweeted"].isnull()] #to check the empty rows
```

Out[0]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_i
19	888202515573088257	NaN	NaN	2017-07-21 01:02:36 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: This is Canela. She attempted s...	8.874740e 17
95	873697596434513921	NaN	NaN	2017-06-11 00:25:14 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: This is Walter. He won't start ...	8.688804e 17
101	872668709062186393	NaN	NaN	2017-06-08	<a	RT @loganamnosis:	8.726576e

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id
	7			04:17:07 +0000	href="http://twitter.com/download/iphone" r...	Penelope here is doing me qu...	17
104	8722617132944957475	NaN	NaN	2017-06-07 01:19:32 +0000	<a href="http://twitter.com/download/iphone" r...	This is Harry. His ears are activated one at a...	NaN
118	8699887020717779329	NaN	NaN	2017-05-31 18:47:24 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: We only rate dogs. This is quit...	8.591970e17
132	8668162802838077444	NaN	NaN	2017-05-23 00:41:20 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: This is Jamesy. He gives a kiss...	8.664507e17
155	8617699731816243220	NaN	NaN	2017-05-09 02:29:07 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: "Good afternoon class today we'...	8.066291e17
182	8566029935878881830	NaN	NaN	2017-04-24 20:17:23 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: This is Luna. It's her first ti...	8.447048e17
211	8519539026226585660	NaN	NaN	2017-04-12 00:23:33 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: This is Astrid. She's a guide d...	8.293743e17
247	8454590767966167055	NaN	NaN	2017-03-25 02:15:26 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: Here's a heartwarming scene of ...	7.562885e17
253	8447047884031139884	NaN	NaN	2017-03-23 00:18:10 +0000	<a href="http://twitter.com/download/iphone" r...	This is Luna. It's her first time outside and ...	NaN
260	8428922088649236448	NaN	NaN	2017-03-18 00:15:37 +0000	<a href="http://twitter.com/download/iphone"	RT @dog_rates: This is Stephan. He just wants ...	8.071068e17

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_i
					<a href="http://twitter.com/download/iphone" r...	ne" r...	
296	837366284874571778	NaN	NaN	2017-03-02 18:17:34 +0000	<a href="http://twitter.com/download/iphone" r...	This is Lucy. She has a portrait of herself on...	NaN
298	837012587749474308	NaN	NaN	2017-03-01 18:52:06 +0000	<a href="http://twitter.com/download/iphone" r...	RT @KennyFromDaBlok: 14/10 h*ckin good hats. w...	8.370113e 17
363	829374341691346946	NaN	NaN	2017-02-08 17:00:26 +0000	<a href="http://twitter.com/download/iphone" r...	This is Astrid. She's a guide doggo in trainin...	NaN
382	827228250799742977	NaN	NaN	2017-02-02 18:52:38 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: This is Phil. He's an important...	6.946697e 17
506	812747805718642688	NaN	NaN	2016-12-24 19:52:31 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: Meet Sammy. At first I was like...	6.800555e 17
566	802247111496568832	NaN	NaN	2016-11-25 20:26:31 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: Everybody drop what you're doin...	7.790561e 17
750	779123168116150273	NaN	NaN	2016-09-23 01:00:13 +0000	<a href="http://twitter.com/download/iphone" r...	This is Reggie. He hugs everyone he meets. 12/...	NaN
784	775096608509886464	NaN	NaN	2016-09-11 22:20:06 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: After so many requests, this is...	7.403732e 17
815	771004394259247104	NaN	NaN	2016-08-31 15:19:06 +0000	<a href="http://twitter.com/download/iphone" r...	RT @katieornah: @dog_rates learning a lot at c...	7.710021e 17
818	77074392396270796	NaN	NaN	2016-08-30	<a	RT @dog_rates:	7.392382e

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id
	8			22:04:05 +0000	href="http://twitter.com/download/iphone" r...	Here's a doggo blowing bubbles....	17
890	759566828574212096	NaN	NaN	2016-07-31 01:50:18 +0000	<a href="http://twitter.com/download/iphone" r...	RT @dog_rates: This... is a Tyrannosaurus rex....	7.395441e 17
932	754011816964026368	NaN	NaN	2016-07-15 17:56:40 +0000	<a href="http://twitter.com/download/iphone" r...	This is Charlie. He pouts until he gets to go ...	NaN
1726	680055455951884288	NaN	NaN	2015-12-24 16:00:30 +0000	<a href="https://about.twitter.com/products/tw...	Meet Sammy. At first I was like "that's a snow...	NaN

In [0]:
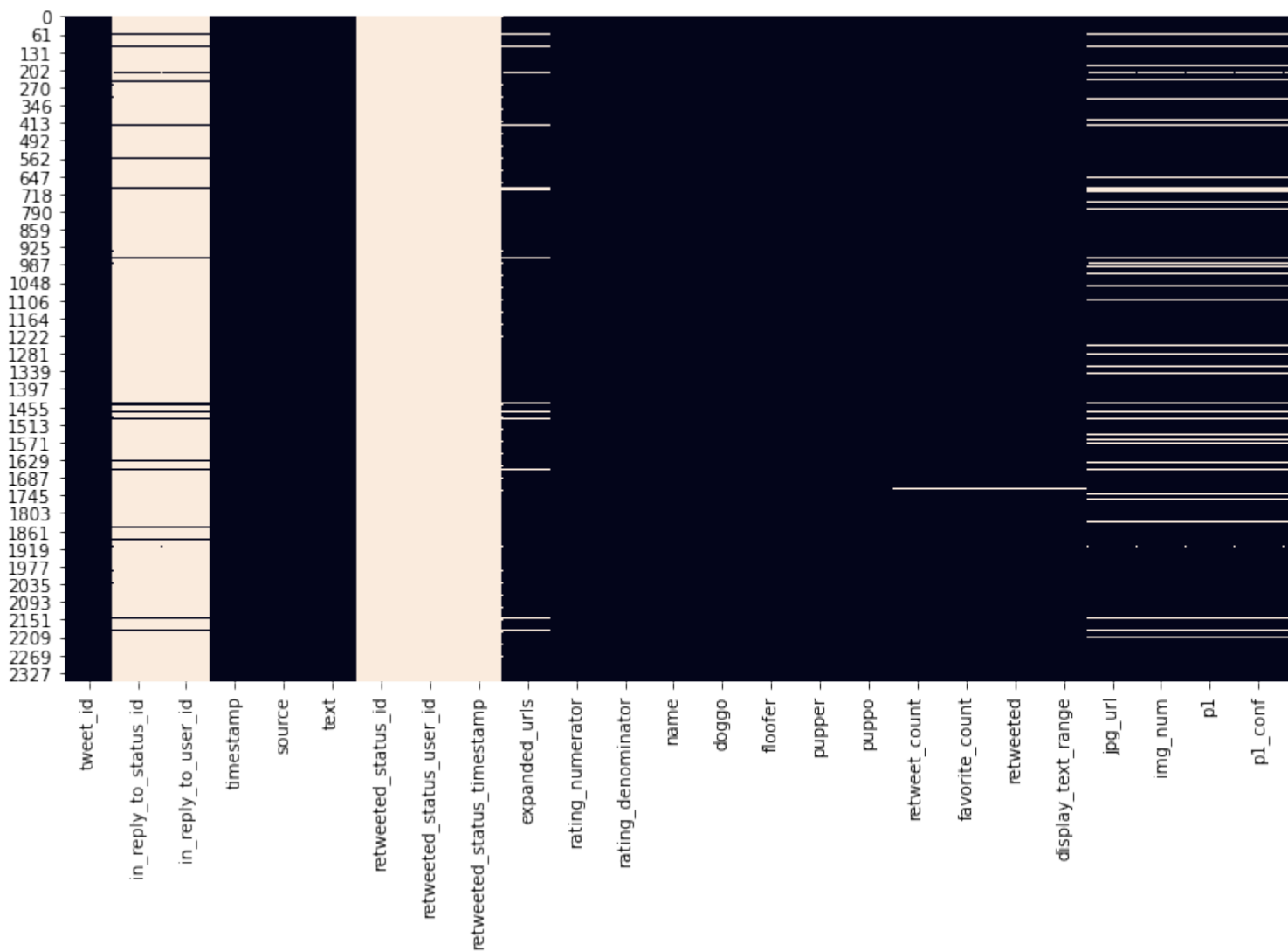
```
#we only want the rows without an entry in "retweeted_status_id" in our master
dataframe
df_master_clean = df_master_clean[df_master_clean["retweeted_status_id"].isnull()]
```

In [0]:

```
fig, ax = plt.subplots(figsize = (20,7))
ax = sns.heatmap(df_master_clean.isnull(), vmin = 0, vmax = 1)
```

In [0]:

```
#check with the column from the api table, no retweets left
df_master_clean.retweeted.value_counts()
```

Out[0]:

```
False    2168
Name: retweeted, dtype: int64
```
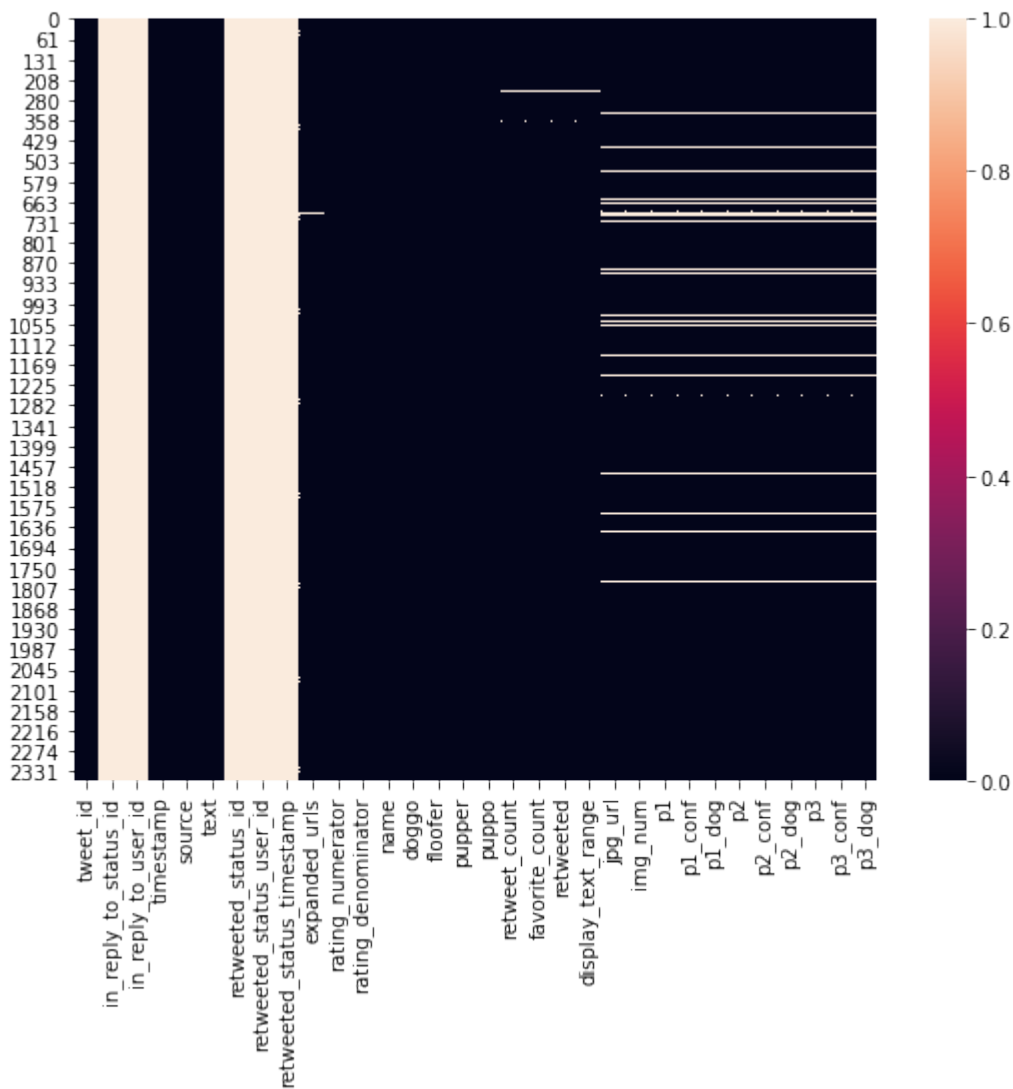
In [0]:

```
#same as for the retweets, we only want the rows without an entry in
"in_reply_to_status_id"
df_master_clean = df_master_clean[df_master_clean.in_reply_to_status_id.isnull()]
```

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_master_clean.isnull(), vmin = 0, vmax = 1)
```

During the gathering process from the API there were some tweets, which got deleted by the account. We will also drop them out of our master dataframe.

In [0]:

```
df_master_clean.dropna(subset = ["retweeted"], inplace = True)
```

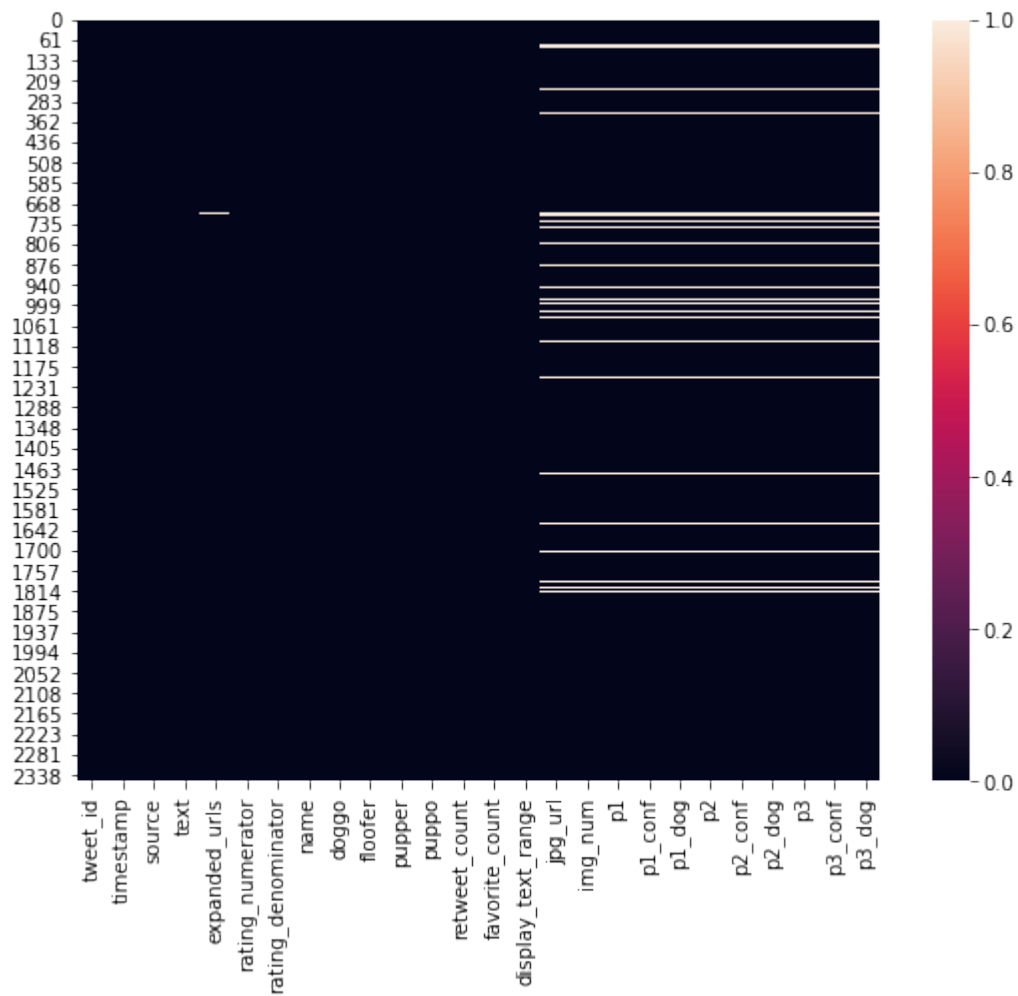In [0]:

```
#drop the unneeded columns
df_master_clean.drop(["in_reply_to_status_id", "in_reply_to_user_id",
                      "retweeted_status_id", "retweeted_status_user_id",
                      "retweeted_status_timestamp", "retweeted"], inplace=True, axis =
1)
```

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_master_clean.isnull(), vmin = 0, vmax = 1)
```

In [0]:

```
df_master_clean.dropna(subset = ["jpg_url"], inplace = True)
```
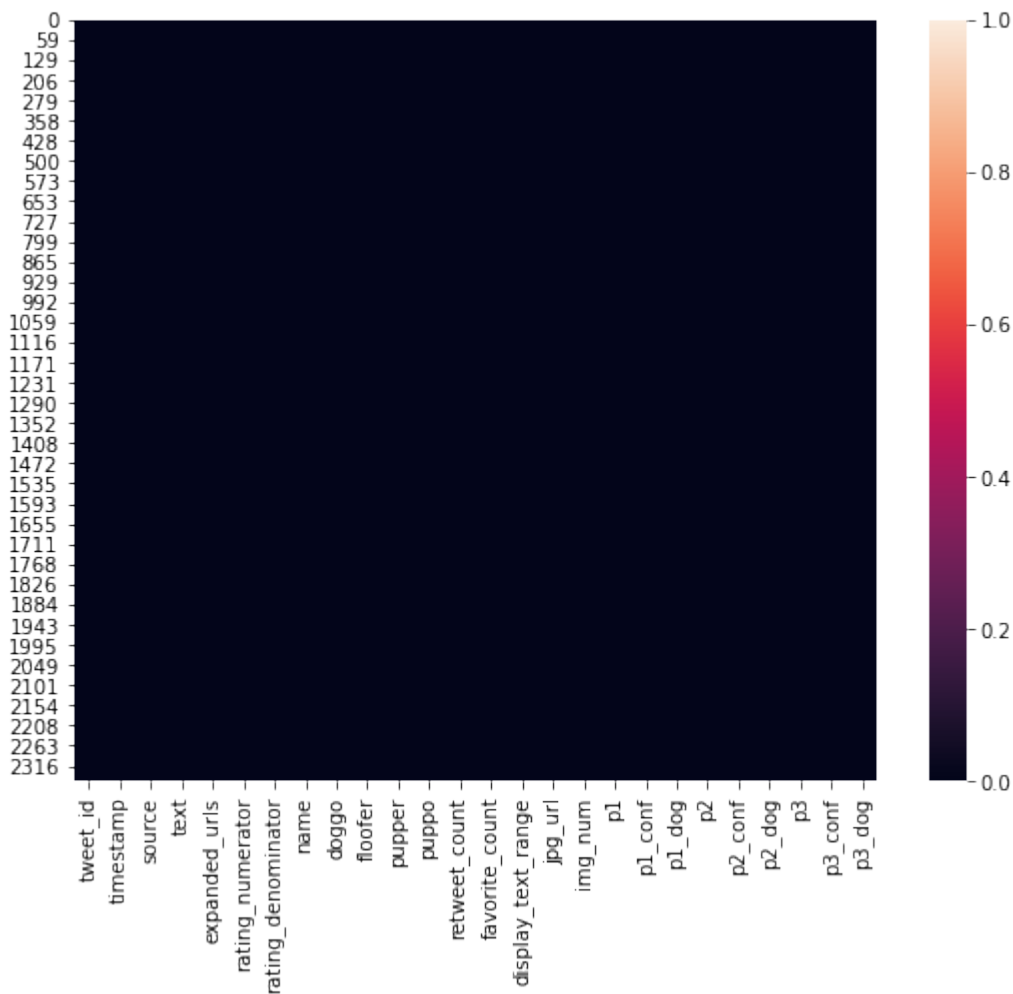
In [0]:

```
sum(df_master_clean.jpg_url.duplicated())
```

Out[0]:

```
0
```

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_master_clean.isnull(), vmin = 0, vmax = 1)
```

Looks like we have cleaned all the missing data

In [0]:

```
#to clean all the images that doesnot displays dog
df_master_clean.drop(df_master_clean.query("p1_dog == False and p2_dog == False and
p3_dog == False").index, inplace = True)
```

In [0]:

```
df_master_clean.query("p1_dog == False and p2_dog == False and p3_dog == False")
```

Out[0]:

tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator	name	doggo	floofer	pupper	pu

Test:

In [0]:

```
df_master_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1659 entries, 1 to 2355
```

```
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   tweet_id           1659 non-null   int64
 1   timestamp          1659 non-null   object
 2   source             1659 non-null   object
 3   text               1659 non-null   object
 4   expanded_urls      1659 non-null   object
 5   rating_numerator   1659 non-null   int64
 6   rating_denominator 1659 non-null   int64
 7   name               1659 non-null   object
 8   doggo              1659 non-null   object
 9   floofer            1659 non-null   object
 10  pupper             1659 non-null   object
 11  puppo              1659 non-null   object
 12  retweet_count      1659 non-null   float64
 13  favorite_count     1659 non-null   float64
 14  display_text_range 1659 non-null   object
 15  jpg_url            1659 non-null   object
 16  img_num            1659 non-null   float64
 17  p1                 1659 non-null   object
 18  p1_conf            1659 non-null   float64
 19  p1_dog             1659 non-null   object
 20  p2                 1659 non-null   object
 21  p2_conf            1659 non-null   float64
 22  p2_dog             1659 non-null   object
 23  p3                 1659 non-null   object
 24  p3_conf            1659 non-null   float64
 25  p3_dog             1659 non-null   object
dtypes: float64(6), int64(3), object(17)
memory usage: 349.9+ KB
```

Define:

Clean the datatypes of the columns

Code:

In [0]:

```
#to change the datatype
df_master_clean["tweet_id"] = df_master_clean["tweet_id"].astype("str")
```

In [0]:

```
#transform the timestamp to datetime
df_master_clean["timestamp"] = pd.to_datetime(df_master_clean.timestamp)
```

In [O]:

```
for x in ["retweet_count", "favorite_count", "img_num"]:
    df_master_clean[x] = df_master_clean[x].astype("int64")
```

Test:

In [O]:

```
df_master_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1659 entries, 1 to 2355
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   tweet_id           1659 non-null   object
 1   timestamp          1659 non-null   datetime64[ns, UTC]
 2   source             1659 non-null   object
 3   text               1659 non-null   object
 4   expanded_urls      1659 non-null   object
 5   rating_numerator   1659 non-null   int64
 6   rating_denominator 1659 non-null   int64
 7   name               1659 non-null   object
 8   doggo              1659 non-null   object
 9   floofer            1659 non-null   object
 10  pupper             1659 non-null   object
 11  puppo              1659 non-null   object
 12  retweet_count      1659 non-null   int64
 13  favorite_count     1659 non-null   int64
 14  display_text_range 1659 non-null   object
 15  jpg_url            1659 non-null   object
 16  img_num            1659 non-null   int64
 17  p1                 1659 non-null   object
 18  p1_conf            1659 non-null   float64
 19  p1_dog             1659 non-null   object
 20  p2                 1659 non-null   object
 21  p2_conf            1659 non-null   float64
 22  p2_dog             1659 non-null   object
 23  p3                 1659 non-null   object
 24  p3_conf            1659 non-null   float64
 25  p3_dog             1659 non-null   object
dtypes: datetime64[ns, UTC](1), float64(3), int64(5), object(17)
memory usage: 349.9+ KB
```

Define:

Clean the wrong numerators - the floats on the one hand, the ones with multiple occurence of the pattern on the other

Now we need to correct all the numerators. And since we have dropped a lot of rows we dont know which problems are still there so lets extract again

Code:

In [0]:

```
pattern = "(\d+\.\d+\/\d+)"

df_master_clean.text.str.extract(pattern, expand = True)[0].dropna()
```

Out[0]:

```
45        13.5/10
695        9.75/10
763      11.27/10
1712     11.26/10
Name: 0, dtype: object
```

In [0]:

```
#get the right numerator out of the string
df_num_clean = df_master_clean.text.str.extract(pattern, expand = True)
[0].dropna().str.split('/', n=1, expand=True)[0]
```

In [0]:

```
df_num_clean
```

Out[0]:

```
45        13.5
695        9.75
763      11.27
1712     11.26
Name: 0, dtype: object
```

In [0]:

```
#get the index of the wrong data
df_num_clean_index = df_num_clean.index
df_num_clean_values = df_num_clean.values.astype("float64")
```

In [0]:

```
#transform the datatypes to float
df_master_clean.rating_numerator = df_master_clean.rating_numerator.astype("float64")
df_master_clean.rating_denominator =
```

```python
df_master_clean.rating_denominator.astype("float64")
#impute the data
df_master_clean.loc[df_num_clean_index, "rating_numerator"] = df_num_clean_values
df_master_clean.loc[df_num_clean_index].rating_numerator
```

Out[0]:

```
45       13.50
695       9.75
763      11.27
1712     11.26
Name: rating_numerator, dtype: float64
```

In [0]:

```python
#There were multiple occurence in ratings so to clean that
pattern = "(\d+(\.\d+)?\/\d+(\.\d+)?)"

print(df_master_clean.text.str.count(pattern)[df_master_clean.text.str.count(pattern) !
= 1])

#get the index of the rows which contains the pattern more than once
pattern_clean_index = df_master_clean.text.str.count(pattern)
[df_master_clean.text.str.count(pattern) != 1].index
```

```
766      2
1007     2
1068     2
1165     2
1202     2
1222     2
1359     2
1465     2
1508     2
1525     2
1538     2
1662     2
1795     2
1832     2
1897     2
1901     2
1970     2
2010     3
2064     2
2113     2
2177     2
```

```
2216    2
2263    2
2272    2
2306    2
2335    2
Name: text, dtype: int64
```

In [0]:

```
df_master_clean.drop(pattern_clean_index, inplace = True)
```

In [0]:

```
df_master_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1633 entries, 1 to 2355
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   tweet_id           1633 non-null   object
 1   timestamp          1633 non-null   datetime64[ns, UTC]
 2   source             1633 non-null   object
 3   text               1633 non-null   object
 4   expanded_urls      1633 non-null   object
 5   rating_numerator   1633 non-null   float64
 6   rating_denominator 1633 non-null   float64
 7   name               1633 non-null   object
 8   doggo              1633 non-null   object
 9   floofer            1633 non-null   object
 10  pupper             1633 non-null   object
 11  puppo              1633 non-null   object
 12  retweet_count      1633 non-null   int64
 13  favorite_count     1633 non-null   int64
 14  display_text_range 1633 non-null   object
 15  jpg_url            1633 non-null   object
 16  img_num            1633 non-null   int64
 17  p1                 1633 non-null   object
 18  p1_conf            1633 non-null   float64
 19  p1_dog             1633 non-null   object
 20  p2                 1633 non-null   object
 21  p2_conf            1633 non-null   float64
 22  p2_dog             1633 non-null   object
 23  p3                 1633 non-null   object
 24  p3_conf            1633 non-null   float64
 25  p3_dog             1633 non-null   object
```

```
dtypes: datetime64[ns, UTC](1), float64(5), int64(3), object(17)
memory usage: 344.5+ KB
```

Test:

In [0]:

```
#no more occurrences of the mentioned problem are left
print(df_master_clean.text.str.count(pattern)[df_master_clean.text.str.count(pattern) !
= 1])
```

```
Series([], Name: text, dtype: int64)
```

Define:

Extract the source from html code

Code:

In [0]:

```
df_master_clean.head(2)
```

Out[0]:

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator	name
1	8921774213063434326	2017-08-01 00:17:27+00:00	<a href="http://twitter.com/download/iphone" r...	This is Tilly. She's just checking pup on you....	https://twitter.com/dog_rates/status/892177421...	13.0	10.0	Tilly
2	8918151813780848644	2017-07-31 00:18:03+00:00	<a href="http://twitter.com/download/iphone" r...	This is Archie. He is a rare Norwegian Pouncin...	https://twitter.com/dog_rates/status/891815181...	12.0	10.0	Archie

we can see that here the image is stored in a html tag so we need to extract it out of the HTML tag

In [0]:

```
df_master_clean.source = df_master_clean.source.str.extract("\>(.*?)\<", expand = True)
```

Test:

In [0]:

```
df_master_clean.iloc[:,:3].head(4)
```

Out[0]:

	tweet_id	timestamp	source
1	89217742130634342 6	2017-08-01 00:17:27+00:00	Twitter for iPhone
2	89181518137808486 4	2017-07-31 00:18:03+00:00	Twitter for iPhone
3	89168955727985868 8	2017-07-30 15:58:51+00:00	Twitter for iPhone
4	89132755892668825 6	2017-07-29 16:00:24+00:00	Twitter for iPhone

In [0]:

```
df_master_clean.source.value_counts()
```

Out[0]:

```
Twitter for iPhone    1605
Twitter Web Client      20
TweetDeck                8
Name: source, dtype: int64
```

Define:

Split the text range into two separate columns

Code:

In [0]:

```
df_master_clean[["display_text_range"]].info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1633 entries, 0 to 1632
Data columns (total 1 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   display_text_range  1633 non-null   object
dtypes: object(1)
memory usage: 12.9+ KB
```

In [0]:

```
df_master_clean.display_text_range[1]
```

Out[0]:

```
[0, 121]
```

In [0]:

```
#get the lower text range at list index 0
df_master_clean["lower_text_range"] =
df_master_clean["display_text_range"].apply(lambda x: x[0])

#get the lower text range at list index 1
df_master_clean["upper_text_range"] =
df_master_clean["display_text_range"].apply(lambda x: x[1])
df_master_clean.drop("display_text_range", axis = 1, inplace = True)
```

Test:

In [0]:

```
df_master_clean[["lower_text_range", "upper_text_range"]].head()
```

Out[0]:

	lower_text_range	upper_text_range
0	0	138
1	0	121
2	0	79
3	0	138
4	0	138

Define:

Remove the "None" out of the doggo, floofer, pupper and puppo column and merge them into one column

Code:

Now we need to merge the doggo,puppo,etc into one column

In [0]:

```
#to merge all the columns at first we need to replace the none value with " "
#replace "None" with "" in each column
for x in ["doggo", "floofer", "pupper", "puppo"]:
    df_master_clean[x].replace("None", "", inplace = True)

#concatenate the columns
df_master_clean['dog_class'] = df_master_clean['doggo'].map(str) + df_master_clean[
    'floofer'].map(str) + df_master_clean['pupper'].map(str) +
df_master_clean['puppo'].map(str)
df_master_clean.dog_class.value_counts()
```

Out[0]:

```
                1379
pupper           164
doggo             53
```

```
puppo               21
floofer              7
doggopupper          7
doggopuppo           1
doggofloofer         1
Name: dog_class, dtype: int64
```

In [O]:

```python
#replace the leftover "" with np.nan
df_master_clean["dog_class"].replace("", np.nan, inplace = True)
```

In [O]:

```python
df_master_clean.dog_class.value_counts()
```

Out[O]:

```
pupper             164
doggo               53
puppo               21
floofer              7
doggopupper          7
doggopuppo           1
doggofloofer         1
Name: dog_class, dtype: int64
```

In [O]:

```python
#count the occurrences of the pattern and show the rows with count > 1
df_master_clean.text.str.count(r'(\bpuppo\b|\bdoggo\b|\bfloofer\b|\bpupper\b)')[
    df_master_clean.text.str.count(r'(\bpuppo\b|\bdoggo\b|\bfloofer\b|\bpupper\b)') >
1]
```

Out[O]:

```
191      2
531      3
575      2
889      2
956      2
1063     2
1113     2
1304     2
1340     2
1367     2
1653     2
1788     2
```

```
1828    2
1907    3
Name: text, dtype: int64
```

In [0]:

```python
df_master_clean["dog_class_re"] = df_master_clean.text.str.extract(
    r'(\bpuppo\b|\bdoggo\b|\bfloofer\b|\bpupper\b)', expand = True)
```

In [0]:

```python
pd.set_option('display.max_colwidth', -1)

#find the differences of the extract
df_master_clean[["text","dog_class", "dog_class_re"]].dropna(subset =
["dog_class_re"]).query("dog_class != dog_class_re")
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: FutureWarning: Passing
a negative integer is deprecated in version 1.0 and will not be supported in future
version. Instead, use None to not limit the column width.
  """Entry point for launching an IPython kernel.
```

Out[0]:

	text	dog_class	dog_class_re
191	Here's a puppo participating in the #ScienceMarch. Cleverly disguising her own doggo agenda. 13/10 would keep the planet habitable for https://t.co/cMhq16isel	doggopuppo	puppo
200	At first I thought this was a shy doggo, but it's actually a Rare Canadian Floofer Owl. Amateurs would confuse the two. 11/10 only send dogs https://t.co/TXdT3tmuYk	doggofloofer	doggo
531	Here we have Burke (pupper) and Dexter (doggo). Pupper wants to be exactly like doggo. Both 12/10 would pet at same time https://t.co/ANBpEYHaho	doggopupper	pupper
575	This is Bones. He's being haunted by another doggo of roughly the same size. 12/10 deep breaths pupper everything's fine https://t.co/55Dqe0SJNj	doggopupper	doggo
889	Meet Maggie & Lila. Maggie is the doggo, Lila is the pupper. They are sisters. Both 12/10 would pet at the same time https://t.co/MYwR4DQKll	doggopupper	doggo
956	Please stop sending it pictures that don't even have a doggo or pupper in them. Churlish af. 5/10 neat couch tho https://t.co/u2c9c7qSg8	doggopupper	doggo
1063	This is just downright precious af. 12/10 for both pupper and doggo https://t.co/o5J479bZUC	doggopupper	pupper
1113	Like father (doggo), like son (pupper). Both 12/10 https://t.co/pG2inLaOda	doggopupper	doggo

In [0]:

```python
#I read the text and corrected the dog class
df_master_clean.loc[191, "dog_class"] = "puppo"
df_master_clean.loc[200, "dog_class"] = "floofer"
df_master_clean.loc[531, "dog_class"] = np.nan
df_master_clean.loc[575, "dog_class"] = "pupper"
df_master_clean.loc[889, "dog_class"] = np.nan
df_master_clean.loc[956, "dog_class"] = np.nan
df_master_clean.loc[1063, "dog_class"] = np.nan
df_master_clean.loc[1113, "dog_class"] = np.nan
```

In [0]:

```python
df_master_clean[["text","dog_class", "dog_class_re"]].dropna(subset =
["dog_class_re"]).query("dog_class != dog_class_re")
```

Out[0]:

	text	dog_class	dog_class_re
200	At first I thought this was a shy doggo, but it's actually a Rare Canadian Floofer Owl. Amateurs would confuse the two. 11/10 only send dogs https://t.co/TXdT3tmuYk	floofer	doggo
531	Here we have Burke (pupper) and Dexter (doggo). Pupper wants to be exactly like doggo. Both 12/10 would pet at same time https://t.co/ANBpEYHaho	NaN	pupper
575	This is Bones. He's being haunted by another doggo of roughly the same size. 12/10 deep breaths pupper everything's fine https://t.co/55Dqe0SJNj	pupper	doggo
889	Meet Maggie & Lila. Maggie is the doggo, Lila is the pupper. They are sisters. Both 12/10 would pet at the same time https://t.co/MYwR4DQKll	NaN	doggo
956	Please stop sending it pictures that don't even have a doggo or pupper in them. Churlish af. 5/10 neat couch tho https://t.co/u2c9c7qSg8	NaN	doggo
1063	This is just downright precious af. 12/10 for both pupper and doggo https://t.co/o5J479bZUC	NaN	pupper
1113	Like father (doggo), like son (pupper). Both 12/10 https://t.co/pG2inLaOda	NaN	doggo

In [0]:

```python
#drop the columns out
df_master_clean.drop(["doggo", "floofer", "pupper", "puppo", "dog_class_re"], inplace =
True, axis = 1)
```

Test:

In [0]:

```python
#count the occurrences of the pattern and show the rows with count > 1
df_master_clean[["text", "dog_class"]].loc[
    df_master_clean.text.str.count(r'(\bpuppo\b|\bdoggo\b|\bfloofer\b|\bpupper\b)')[
```

```
      df_master_clean.text.str.count(r'(\bpuppo\b|\bdoggo\b|\bfloofer\b|\bpupper\b)')
> 1].index]
```

Out[0]:

	text	dog_class
191	Here's a puppo participating in the #ScienceMarch. Cleverly disguising her own doggo agenda. 13/10 would keep the planet habitable for https://t.co/cMhq16isel	puppo
531	Here we have Burke (pupper) and Dexter (doggo). Pupper wants to be exactly like doggo. Both 12/10 would pet at same time https://t.co/ANBpEYHaho	NaN
575	This is Bones. He's being haunted by another doggo of roughly the same size. 12/10 deep breaths pupper everything's fine https://t.co/55Dqe0SJNj	pupper
889	Meet Maggie & Lila. Maggie is the doggo, Lila is the pupper. They are sisters. Both 12/10 would pet at the same time https://t.co/MYwR4DQKll	NaN
956	Please stop sending it pictures that don't even have a doggo or pupper in them. Churlish af. 5/10 neat couch tho https://t.co/u2c9c7qSg8	NaN
1063	This is just downright precious af. 12/10 for both pupper and doggo https://t.co/o5J479bZUC	NaN
1113	Like father (doggo), like son (pupper). Both 12/10 https://t.co/pG2inLaOda	NaN
1304	"I shall trip the big pupper with leash. Big pupper will never see it coming. I am a genius." Both 11/10 https://t.co/uQsCJ8pf51	pupper
1340	Here is a heartbreaking scene of an incredible pupper being laid to rest. 10/10 RIP pupper https://t.co/81mvJ0rGRu	pupper
1367	This is Sansa. She's gotten too big for her chair. Not so smol anymore. 11/10 once a pupper, always a pupper https://t.co/IpAoztle2s	pupper
1653	"Hello forest pupper I am house pupper welcome to my abode" (8/10 for both) https://t.co/qFD8217fUT	pupper
1788	Reckless pupper here. Not even looking at road. Absolute menace. No regard for fellow pupper lives. 10/10 still cute https://t.co/96IBkOYB7j	pupper
1828	All this pupper wanted to do was go skiing. No one told him about the El Niño. Poor pupper. 10/10 maybe next year https://t.co/fTgbq1UBR9	pupper
1907	This pupper just wants a belly rub. This pupper has nothing to do w the tree being sideways now. 10/10 good pupper https://t.co/AyJ7Ohk71f	pupper

All looks correct

Define:

Remove the wrong names of name column

Code:

In [0]:

```
#here we will remove all the wrong names
```

```
for x in ["None", "a", "by", "the"]:
    df_master_clean["name"].replace(x, np.nan, inplace = True)
```

In [0]:

```
#to convert all the lowercase into uppercase
df["name"]= df["name"].str.upper().str.title()
```

Test:

In [165]:

```
df_master_clean.name.value_counts()
```

Out[165]:

```
Cooper     10
Lucy        9
Charlie     9
Tucker      9
Oliver      9
           ..
Jimison     1
Laela       1
Carll       1
Beckham     1
Ralphé      1
Name: name, Length: 833, dtype: int64
```
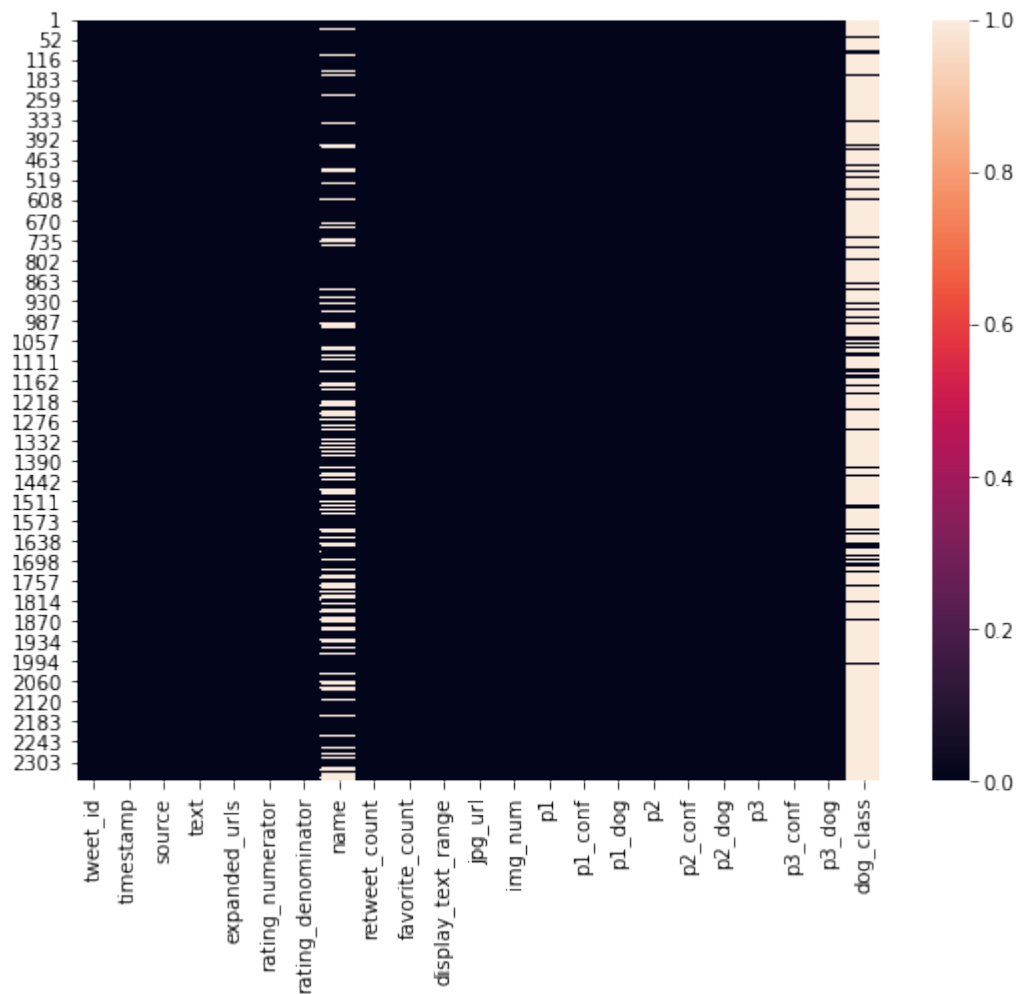
In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_master_clean.isnull(), vmin = 0, vmax = 1)
```

Define:

Reduce the prediction columns into two - breed and conf

Code:

In [0]:

```
#now we will reduce the columns
def get_attr(x):
    """
    INPUT:
        Columns in this order: Check1, Result1, Check2, Result2, Result3
    OUTPUT:
        Results based on the check in this columns
    """
    if x[0] == True:
        return x[1]
    elif x[2] == True:
        return x[3]
    else:
        return x[4]
```

```
df_master_clean["breed"] = df_master_clean[["p1_dog", "p1", "p2_dog", "p2",
"p3"]].apply(get_attr, axis = 1)
df_master_clean["conf"] = df_master_clean[["p1_dog", "p1_conf", "p2_dog", "p2_conf",
"p3_conf"]].apply(get_attr, axis = 1)
```

In [0]:

```
df_master_clean.iloc[:, 12:]
```

Out[0]:

	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog	dog_class	
1	1	Chihuahua	0.323581	True	Pekinese	0.090647	True	papillon	0.068957	True	NaN	Chih
2	1	Chihuahua	0.716012	True	malamute	0.078253	True	kelpie	0.031379	True	NaN	Chih
3	1	paper_towel	0.170278	False	Labrador_retriever	0.168086	True	spatula	0.040836	False	NaN	Labr ieve
4	2	basset	0.555712	True	English_springer	0.225770	True	German_short-haired_pointer	0.175219	True	NaN	bass
5	1	Chesapeake_Bay_retriever	0.425595	True	Irish_terrier	0.116317	True	Indian_elephant	0.076902	False	NaN	Che Bay
...	...	...	...	...	...	...	...	...	...	...	...	...
2351	1	miniature_pinscher	0.560311	True	Rottweiler	0.243682	True	Doberman	0.154629	True	NaN	min sche
2352	1	Rhodesian_ridgeback	0.408143	True	redbone	0.360687	True	miniature_pinscher	0.222752	True	NaN	Rho dgel
2353	1	German_shepherd	0.596461	True	malinois	0.138584	True	bloodhound	0.116197	True	NaN	Gerr herc
2354	1	redbone	0.506826	True	miniature_pinscher	0.074192	True	Rhodesian_ridgeback	0.072010	True	NaN	redb
2355	1	Welsh_springer_spaniel	0.465074	True	collie	0.156665	True	Shetland_sheepdog	0.061428	True	NaN	Wels er_s

1633 rows × 13 columns

In [0]:

```
#drop the reduced columns
df_master_clean.drop(df_master_clean.columns[12:21], inplace = True, axis = 1)
```

Test:

In [0]:

```
df_master_clean.head()
```

Out[0]:

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator
1	892177421306343426	2017-08-01 00:17:27+00:00	Twitter for iPhone	This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she's available for pats, snugs, boops, the whole bit. 13/10 https://t.co/0Xxu71qeIV	https://twitter.com/dog_rates/status/892177421306343426/photo/1	13.0	10.0
2	891815181378084864	2017-07-31 00:18:03+00:00	Twitter for iPhone	This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike. 12/10 https://t.co/wUnZnhtVJB	https://twitter.com/dog_rates/status/891815181378084864/photo/1	12.0	10.0
3	891689557279858688	2017-07-30 15:58:51+00:00	Twitter for iPhone	This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us https://t.co/tD36da7qLQ	https://twitter.com/dog_rates/status/891689557279858688/photo/1	13.0	10.0
4	891327558926688256	2017-07-29 16:00:24+00:00	Twitter for iPhone	This is Franklin. He would like you to stop	https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://	12.0	10.0

	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator
				calling him "cute." He is a very fierce shark and should be respected as such. 12/10 #BarkWeek https://t.co/AtUZn91f7f	twitter.com/dog_rates/status/891327558926688256/photo/1		
5	891087950875897856	2017-07-29 00:08:17+00:00	Twitter for iPhone	Here we have a majestic great white breaching off South Africa's coast. Absolutely h*ckin breathtaking. 13/10 (IG: tucker_marlo) #BarkWeek https://t.co/kQ04fDDRmh	https://twitter.com/dog_rates/status/891087950875897856/photo/1	13.0	10.0

Define:

Clean the new breed column by replacing the "_" with a whitespace and make them all lowercase

Test:

In [0]:

```
#replace "_" with " "
df_master_clean.breed = df_master_clean.breed.str.replace("_", " ")
```

In [0]:

```
df_master_clean.breed
```

Out[0]:

```
1        Chihuahua
2        Chihuahua
3        Labrador retriever
4        basset
```

```
5        Chesapeake Bay retriever
                    ...
2351     miniature pinscher
2352     Rhodesian ridgeback
2353     German shepherd
2354     redbone
2355     Welsh springer spaniel
Name: breed, Length: 1633, dtype: object
```

Test:

In [0]:

```
df_master_clean.breed = df_master_clean.breed.str.lower()

df_master_clean.breed.value_counts().head(10)
```

Out[0]:

```
golden retriever      154
labrador retriever    103
pembroke              93
chihuahua             87
pug                   62
toy poodle            49
chow                  48
samoyed               41
pomeranian            39
malamute              33
Name: breed, dtype: int64
```
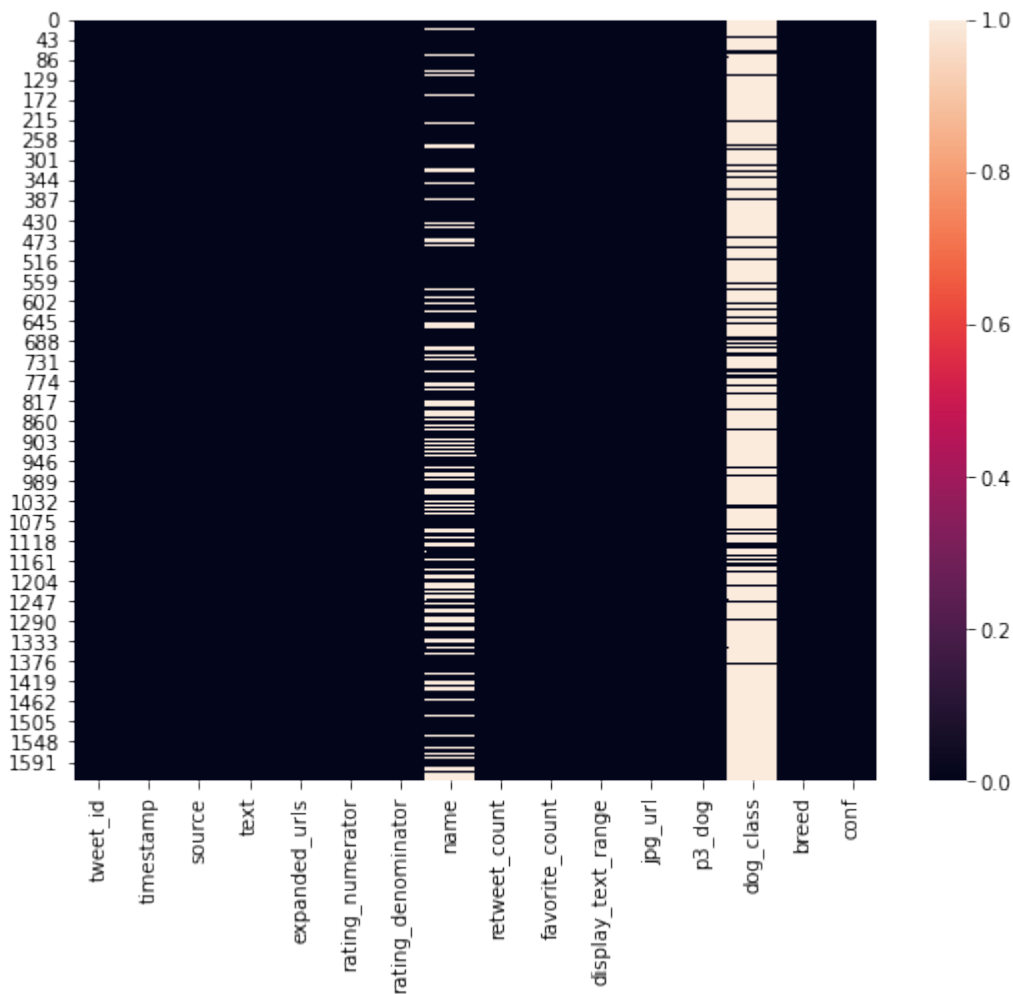
In [0]:

```
#reset index to match with the real amount of rows
df_master_clean.reset_index(drop = True, inplace = True)
```

In [0]:

```
fig, ax = plt.subplots(figsize = (9,7))
ax = sns.heatmap(df_master_clean.isnull(), vmin = 0, vmax = 1)
```

Here we can see that our data is much cleaner now

In [0]:

```
#save the data to a *.csv file
df_master_clean.to_csv('twitter_archive_master.csv', index = False)
```

Data Analysis¶

Questions:

```
1. Based on the predicted, most likely dog breed: Which breed gets retweeted and
favorited the most overall?
2. How did the account develop (speaking about number of tweets, retweets, favorites,
image number and length of the tweets)?
3. Is there a pattern visible in the timing of the tweets?
```

In [0]:

```
#read in the master csv
df = pd.read_csv("twitter_archive_master.csv")
df.head()
```

Out[0]:

|	tweet_id	timestamp	source	text	expanded_urls	rating_numerator	rating_denominator
0	892177421306343426	2017-08-01 00:17:27+00:00	Twitter for iPhone	This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she's available for pats, snugs, boops, the whole bit. 13/10 https://t.co/0Xxu71qeIV	https://twitter.com/dog_rates/status/892177421306343426/photo/1	13.0	10.0
1	891815181378084864	2017-07-31 00:18:03+00:00	Twitter for iPhone	This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike. 12/10 https://t.co/wUnZnhtVJB	https://twitter.com/dog_rates/status/891815181378084864/photo/1	12.0	10.0
2	891689557279858688	2017-07-30 15:58:51+00:00	Twitter for iPhone	This is Darla. She commenced a snooze mid meal. 13/10 happens to the best of us https://t.co/tD36da7qLQ	https://twitter.com/dog_rates/status/891689557279858688/photo/1	13.0	10.0
3	891327558926688256	2017-07-29 16:00:24+00:00	Twitter for iPhone	This is Franklin. He would like you to stop	https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://	12.0	10.0

	tweet_id	timestamp	source	text	expanded_urls	rating_nu merator	rating_deno minator
				calling him "cute." He is a very fierce shark and should be respected as such. 12/10 #BarkWeek https://t.co/At UZn91f7f	twitter.com/dog_rates/ status/ 891327558926688256 /photo/1		
4	89108795087589785 6	2017-07-29 00:08:17+0 0:00	Twitter for iPhone	Here we have a majestic great white breaching off South Africa's coast. Absolutely h*ckin breathtaking. 13/10 (IG: tucker_marlo) #BarkWeek https://t.co/k QO4fDDRmh	https://twitter.com/ dog_rates/status/ 891087950875897856 /photo/1	13.0	10.0

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.countplot(x = "breed", data = df,
order=df.breed.value_counts().iloc[:10].index, palette = "viridis")
ax.set_title("count of classified breeds in the dataset");

ax.set_ylim(0, 170)
#https://stackoverflow.com/questions/33179122/seaborn-countplot-with-frequencies
for p in ax.patches:
    ax.annotate('{:.0f}'.format(p.get_height()), (p.get_x()+0.3, p.get_height()+2))
```

count of classified breeds in the dataset
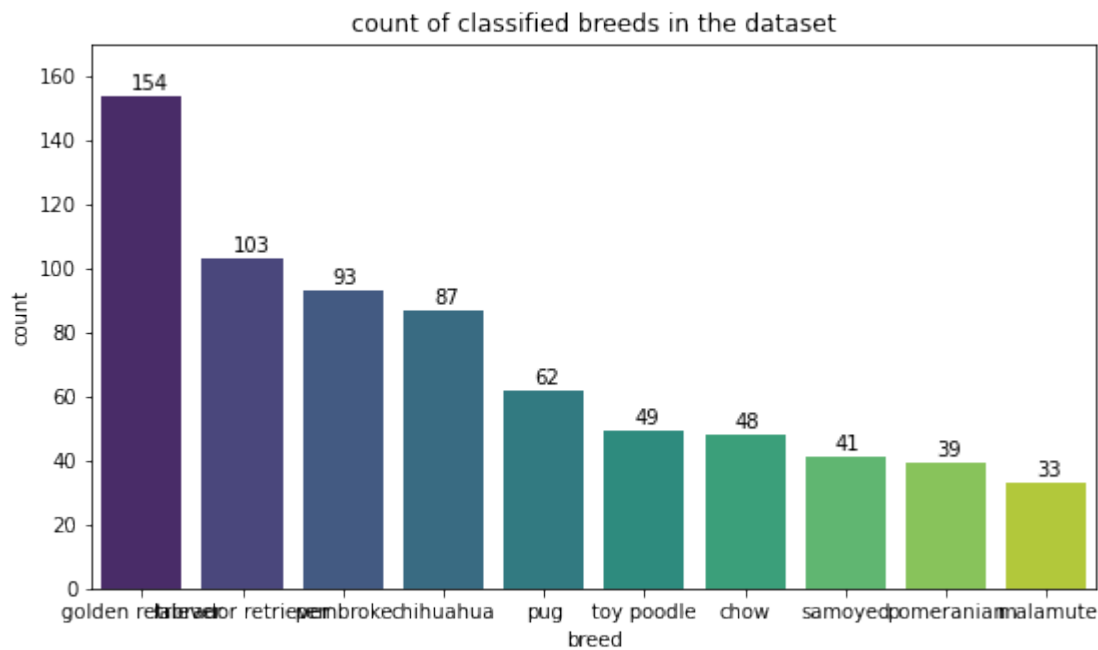
So, we can see that the dogs mostly reviewd is golden retriever and the second is labrador

In [0]:

```
df_breed_group = df[["retweet_count", "favorite_count", "breed"]].groupby("breed",
as_index = False).sum()
```

In [0]:

```
df_breed_group.sort_values("retweet_count", ascending = False).head(10)
```

Out[0]:

	breed	retweet_count	favorite_count
45	golden retriever	505839	1765712
62	labrador retriever	347651	1123149
80	pembroke	258538	969622
27	chihuahua	213031	679620
88	samoyed	170619	517398
41	french bulldog	144684	564115
30	cocker spaniel	127336	368344
28	chow	117604	419155
82	pug	103755	350579
81	pomeranian	99768	294519

In [0]:

```
df_breed_group.sort_values("favorite_count", ascending = False).head(10)
```

Out[0]:

	breed	retweet_count	favorite_count
45	golden retriever	505839	1765712
62	labrador retriever	347651	1123149
80	pembroke	258538	969622
27	chihuahua	213031	679620
41	french bulldog	144684	564115
88	samoyed	170619	517398
28	chow	117604	419155
30	cocker spaniel	127336	368344
82	pug	103755	350579
66	malamute	96978	327871

Here also we can see that the golden retriever and the labrader tops the list

In [0]:

```
df_breed_group["sum"] = df_breed_group["retweet_count"] +
df_breed_group["favorite_count"]


df_breed_group.sort_values("sum", ascending = False).head(10)
```
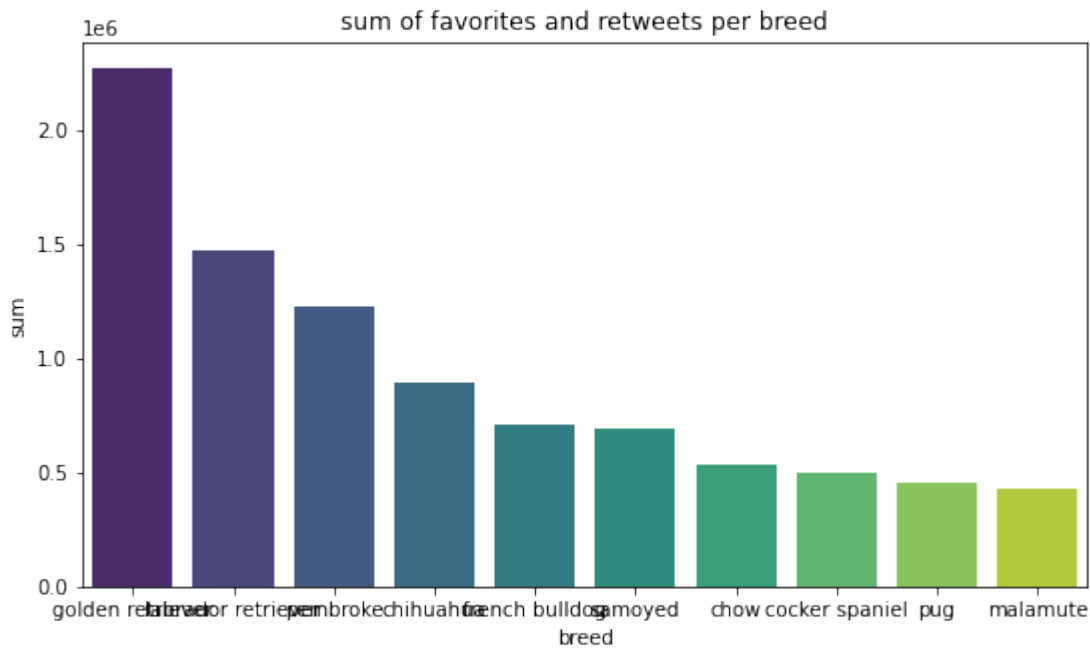
Out[0]:

	breed	retweet_count	favorite_count	sum
45	golden retriever	505839	1765712	2271551
62	labrador retriever	347651	1123149	1470800
80	pembroke	258538	969622	1228160
27	chihuahua	213031	679620	892651
41	french bulldog	144684	564115	708799
88	samoyed	170619	517398	688017
28	chow	117604	419155	536759
30	cocker spaniel	127336	368344	495680
82	pug	103755	350579	454334
66	malamute	96978	327871	424849

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.barplot(x = "breed", y = "sum", data = df_breed_group.sort_values("sum",
ascending=False).iloc[:10], palette = "viridis")
ax.set_title("sum of favorites and retweets per breed");
```

sum of favorites and retweets per breed

In [0]:

```
df[["retweet_count", "favorite_count", "breed"]].sort_values("retweet_count", ascending
= False).head(5)
```

Out[0]:

	retweet_count	favorite_count	breed
678	77999	156520	labrador retriever
705	57820	115635	eskimo dog
356	56865	120909	chihuahua
271	43938	133322	lakeland terrier
53	40676	99428	english springer

In [0]:

```
df[["retweet_count", "favorite_count", "breed"]].sort_values("favorite_count",
ascending = False).head(5)
```

Out[0]:

|	retweet_count	favorite_count	breed
678	77999	156520	labrador retriever
271	43938	133322	lakeland terrier
356	56865	120909	chihuahua
99	33033	116636	french bulldog
705	57820	115635	eskimo dog

In [0]:

```
#see the mean average rating
df_breed_group_mean = df[["rating_numerator", "breed"]].groupby("breed", as_index =
```

```
False).mean()
```

In [0]:

```
fig, ax = plt.subplots(figsize = (7,5))
ax = sns.barplot(x = "breed", y = "rating_numerator", data =
df_breed_group_mean.sort_values("rating_numerator", ascending = False).iloc[:100],
palette = "viridis")
ax.set_title("average rating_numerator per breed");
ax.xaxis.set_ticklabels([])
plt.tight_layout()
```


average rating_numerator per breed
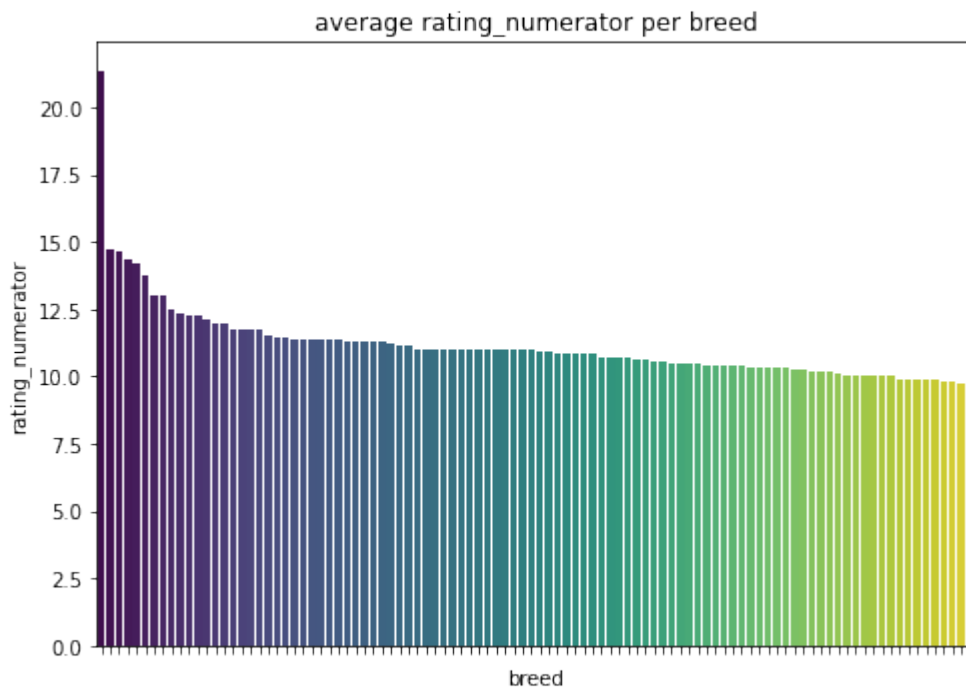
In [0]:

```
df_breed_group_mean.sort_values("rating_numerator", ascending = False).head(10)
```

Out[0]:

	breed	rating_numerator
96	soft-coated wheaten terrier	21.357143
109	west highland white terrier	14.687500
48	great pyrenees	14.666667
16	borzoi	14.333333
28	chow	14.166667
62	labrador retriever	13.728155
4	australian terrier	13.000000
18	bouvier des flandres	13.000000
87	saluki	12.500000

	breed	rating_numerator
21	briard	12.333333

In [0]:

```
df[["breed", "rating_numerator"]].sort_values("rating_numerator", ascending =
False).head(5)
```

Out[0]:

	breed	rating_numerator
583	labrador retriever	165.0
1222	chow	144.0
1116	soft-coated wheaten terrier	121.0
823	golden retriever	99.0
1269	labrador retriever	88.0

In [0]:

```
df.query("breed != 'soft-coated wheaten terrier'").rating_numerator.mean()
```

Out[0]:

11.261754169240271

The soft-coated wheaten terrier got a very high mean rating. In fact, the labrador retriever got overall the biggest rating with 165, but since there are a lot more tweets with labrador retriever than for the soft-coated wheaten terrier, the one big rating of the soft-coated wheaten terrier has a higher weight then the one of the labrador retriever (14 tweets of soft-coated wheaten terrier and 105 of labrador retriever).

Not taking the outlier into account, this leads us to an average rating of 11. Based on the number of posts, retweets, favorites and mean rating, we will give the title of "Most overall liked dog of this Twitter account and its community" to the labrador retriever.

In [0]:

```
df_time = df.copy()
df_time.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1633 entries, 0 to 1632
Data columns (total 16 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   tweet_id            1633 non-null   int64
 1   timestamp           1633 non-null   object
 2   source              1633 non-null   object
 3   text                1633 non-null   object
 4   expanded_urls       1633 non-null   object
```

```
 5   rating_numerator      1633 non-null   float64
 6   rating_denominator    1633 non-null   float64
 7   name                  1190 non-null   object
 8   retweet_count         1633 non-null   int64
 9   favorite_count        1633 non-null   int64
 10  display_text_range    1633 non-null   object
 11  jpg_url               1633 non-null   object
 12  p3_dog                1633 non-null   bool
 13  dog_class             249 non-null    object
 14  breed                 1633 non-null   object
 15  conf                  1633 non-null   float64
dtypes: bool(1), float64(3), int64(3), object(9)
memory usage: 193.1+ KB
```

In [0]:

```python
df_time.timestamp = pd.to_datetime(df_time.timestamp)
```

In [0]:

```python
df_time["dow"] = df_time["timestamp"].apply(lambda x: x.dayofweek)
```

In [0]:

```python
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.countplot(x = "dow", data = df_time, palette = "viridis")
ax.set_title("count of tweets in the dataset for each day of the week");

ax.set_ylim(0, 300)
#https://stackoverflow.com/questions/33179122/seaborn-countplot-with-frequencies
for p in ax.patches:
    ax.annotate('{:.0f}'.format(p.get_height()), (p.get_x()+0.3, p.get_height()+5))
```
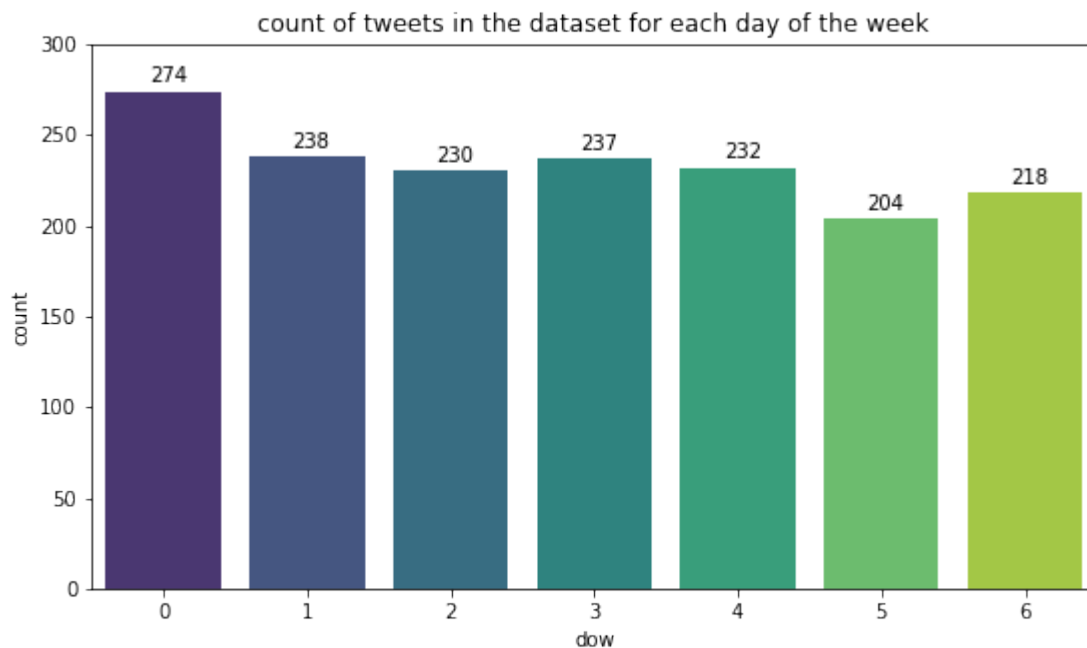
count of tweets in the dataset for each day of the week

we can see that the most tweets are posted on Monday. For tuesday to friday it is nearly the same number of posts. On the weekend the Twitter profile tweets a little bil less

In [0]:

```
df_time.timestamp.min()
```

Out[0]:

```
Timestamp('2015-11-15 22:32:08+0000', tz='UTC')
```

In [0]:

```
df_time.timestamp.max()
```

Out[0]:

```
Timestamp('2017-08-01 00:17:27+0000', tz='UTC')
```

The dataset contains data from 2015 to 2017

In [0]:

```
#get the month out of the timestamp
df_time["month"] = df_time["timestamp"].apply(lambda x: x.month)
#get the year out of the timestamp
df_time["year"] = df_time["timestamp"].apply(lambda x: x.year)
#get the hour out of the timestamp
df_time["hour"] = df_time["timestamp"].apply(lambda x: x.hour)
```
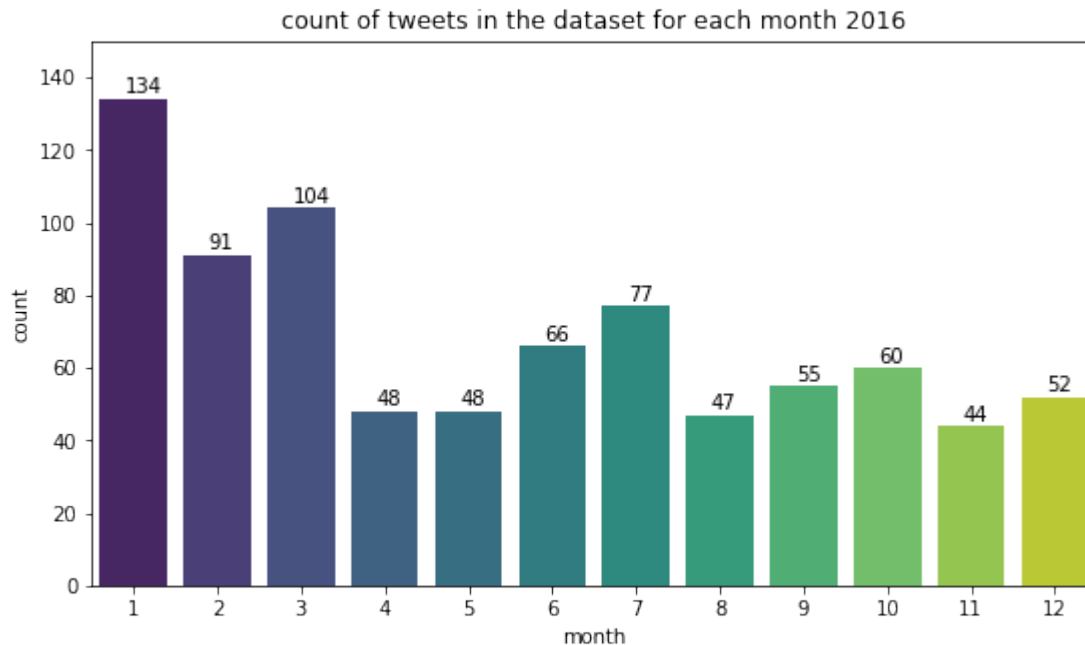
For the first graph I only want to take a look on the full year 2016

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
```

```
ax = sns.countplot(x = "month", data = df_time.query("year == 2016"), palette =
"viridis")
ax.set_title("count of tweets in the dataset for each month 2016");
```

```
ax.set_ylim(0,150)
for p in ax.patches:
    ax.annotate('{:.0f}'.format(p.get_height()), (p.get_x()+0.3, p.get_height()+1.5))
```


Over the timeperiod of 2016 the number of post per months decreased. It went from 134 tweets in January to 52 in December. Does this mean, that the performance of this account is also decreasing?

In [0]:

```
df_time['month_year'] = pd.to_datetime(df["timestamp"]).dt.to_period('M')
```

```
/usr/local/lib/python3.6/dist-packages/pandas/core/arrays/datetimes.py:1104:
UserWarning: Converting to PeriodArray/Index representation will drop timezone
information.
    UserWarning,
```

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.countplot(x = "month_year", data = df_time.sort_values("month_year"), palette
= "viridis", )
ax.set_title("count of tweets in the dataset for each year - month combination in this
dataset");
```
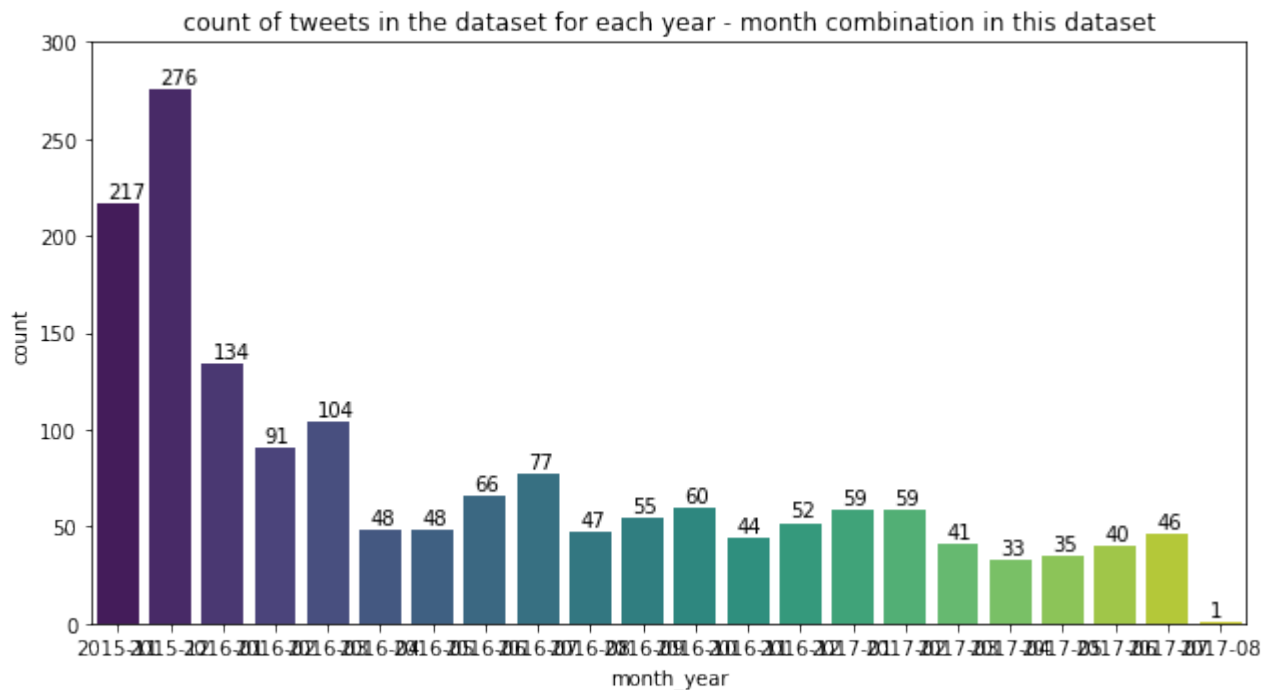
```
ax.set_ylim(0, 300)
```

```
for p in ax.patches:
    ax.annotate('{:.0f}'.format(p.get_height()), (p.get_x()+0.2, p.get_height()+3))
```

```
plt.tight_layout()
```


count of tweets in the dataset for each year - month combination in this dataset

If we look at it over the whole timeperiod it becomes even more clear. In April 2016 the number of tweets dropped and since then it has a relatively stable level. To see if the performance of the Account decreased we will take a look on the favorites and retweets that the posts get.

In [0]:

```
df_time_groupby = df_time.groupby("month_year", as_index = False).sum()
```

In [0]:

```
df_time_groupby.head()
```

Out[0]:

	month_year	tweet_id	rating_nu merator	rating_deno minator	retweet_ count	favorite_ count	p3_d og	conf	dow	month	y
0	2015-11	1.450925e+ 20	2049.00	2170.0	92890.0	237891.0	187.0	105.165214	635.0	2387.0	437
1	2015-12	1.866097e+ 20	2986.26	2940.0	390890.0	962589.0	237.0	143.884391	783.0	3312.0	556
2	2016-01	9.218872e+ 19	1505.00	1440.0	156315.0	446701.0	112.0	71.353070	418.0	134.0	270
3	2016-02	6.362919e+ 19	1033.00	980.0	114418.0	331954.0	80.0	49.415261	256.0	182.0	183

|	month_year	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count	p3_dog	conf	dow	month	y
4	2016-03	7.376839e+19	1298.00	1230.0	141137.0	424219.0	93.0	62.738141	310.0	312.0	209

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.barplot(x = "month_year", y = "favorite_count", data = df_time_groupby,
palette = "viridis")
ax.set_title("sum of favorites per month-year combination");
plt.tight_layout()
```
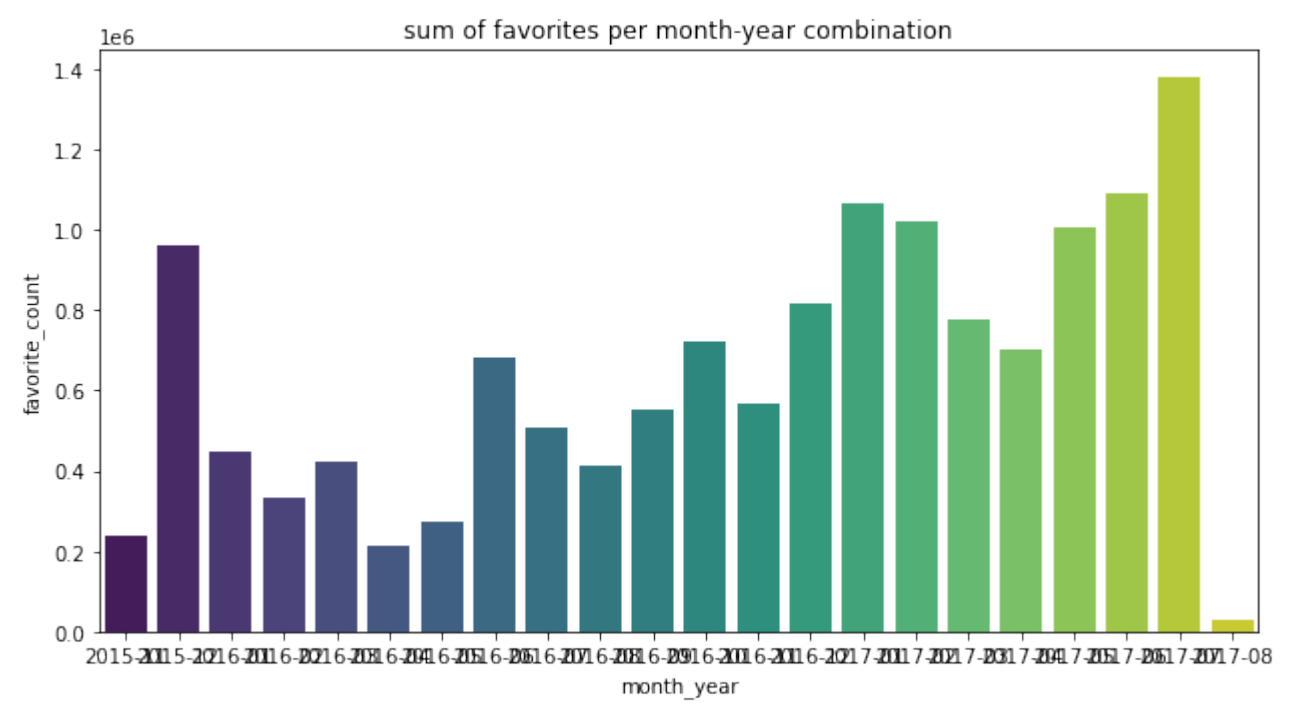

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.barplot(x = "month_year", y = "retweet_count", data = df_time_groupby, palette
= "viridis")
ax.set_title("sum of retweets per month-year combination");
```

sum of retweets per month-year combination
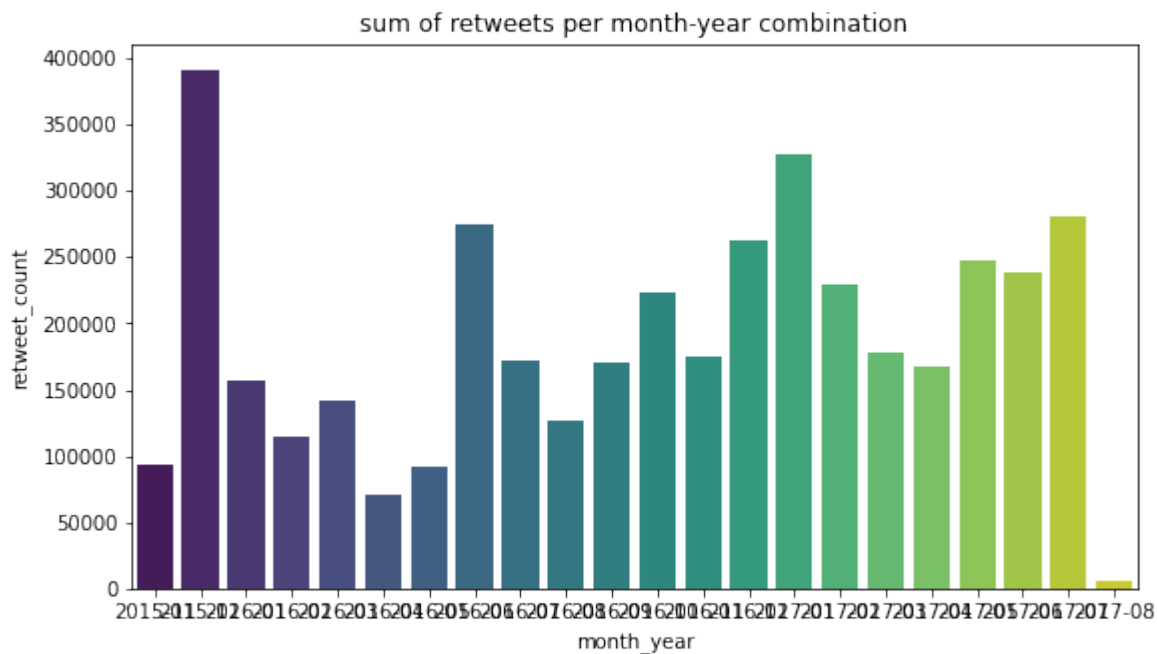
Interesting, while the number of tweets per month is decreasing, the favorites and retweets per month are increasing

In [0]:

```
#create a groupby object over the mean
df_time_groupby_mean = df_time.groupby("month_year", as_index = False).mean()
```
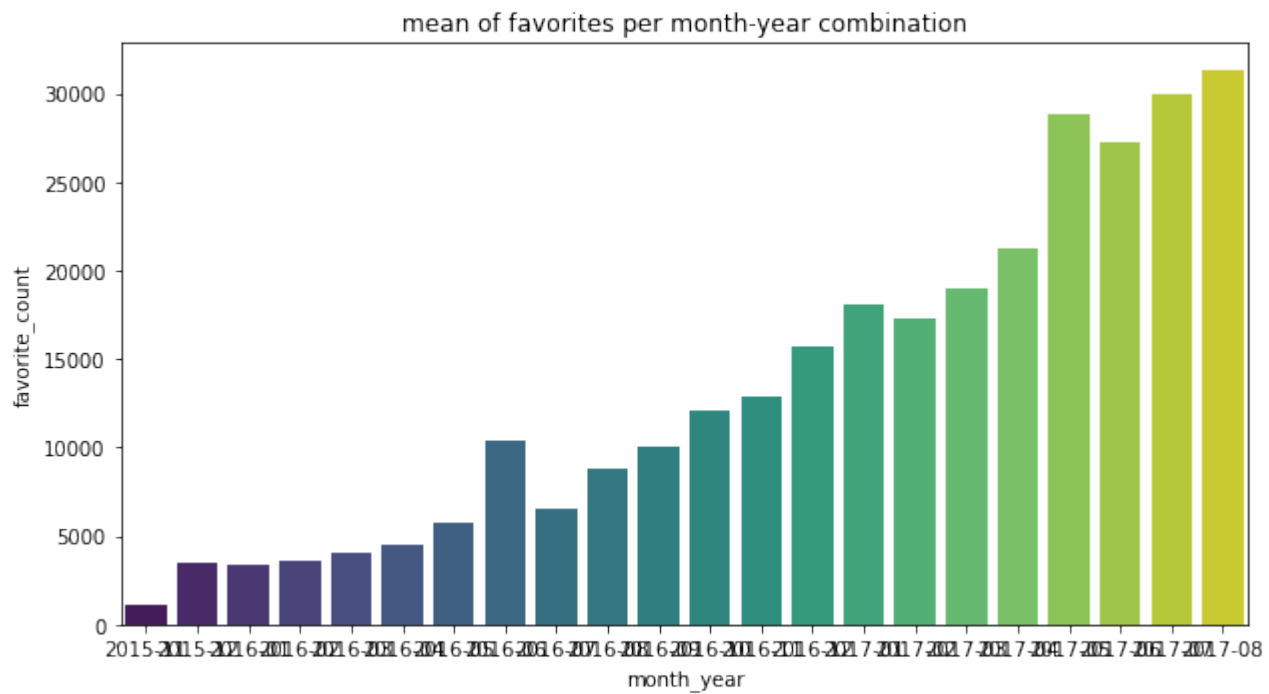
In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.barplot(x = "month_year", y = "favorite_count", data = df_time_groupby_mean,
palette = "viridis")
ax.set_title("mean of favorites per month-year combination");
plt.tight_layout()
```
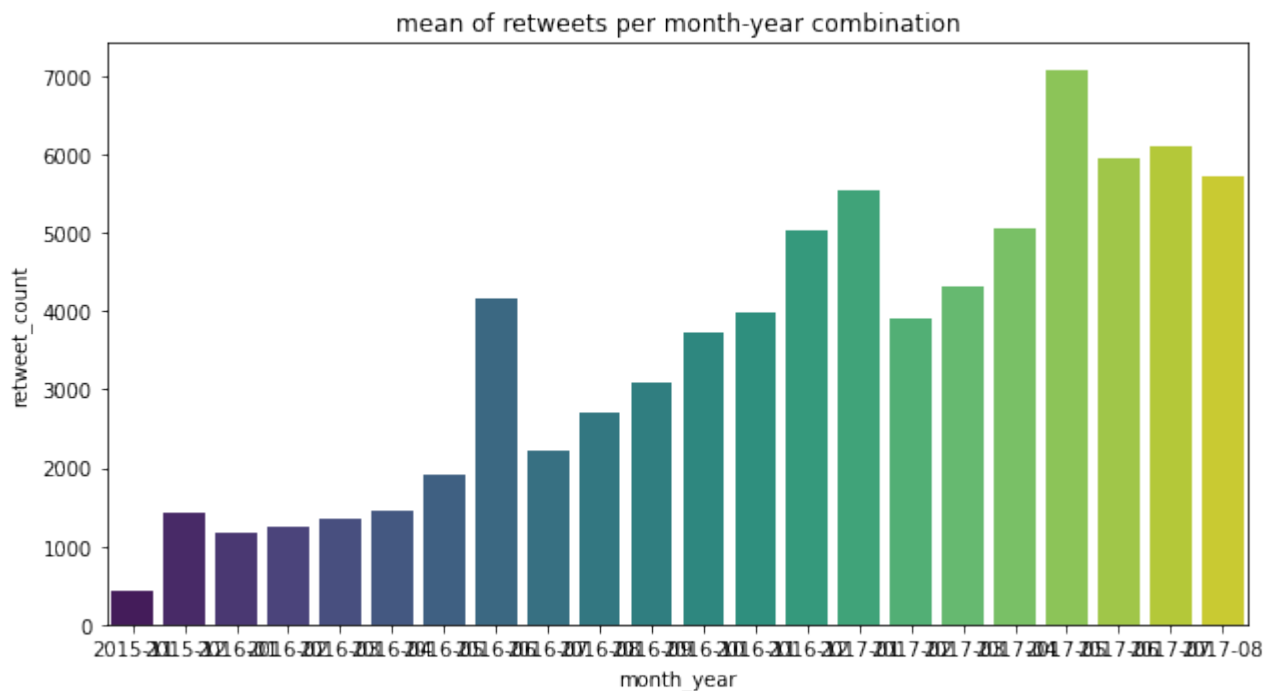
mean of favorites per month-year combination

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.barplot(x = "month_year", y = "retweet_count", data = df_time_groupby_mean,
palette = "viridis")
ax.set_title("mean of retweets per month-year combination");
plt.tight_layout()
```


mean of retweets per month-year combination

If we look at the average number of favorites and retweets the clear uptrend gets even more clearclearer! Now let's see if the number of posted images per month or the average upper text range changed over time.

In [0]:

```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.countplot(x = "hour", data = df_time.query("year == 2016"), palette =
"viridis")
ax.set_title("count of tweets in the dataset for each hour in 2016");
```

count of tweets in the dataset for each hour in 2016

In [0]:
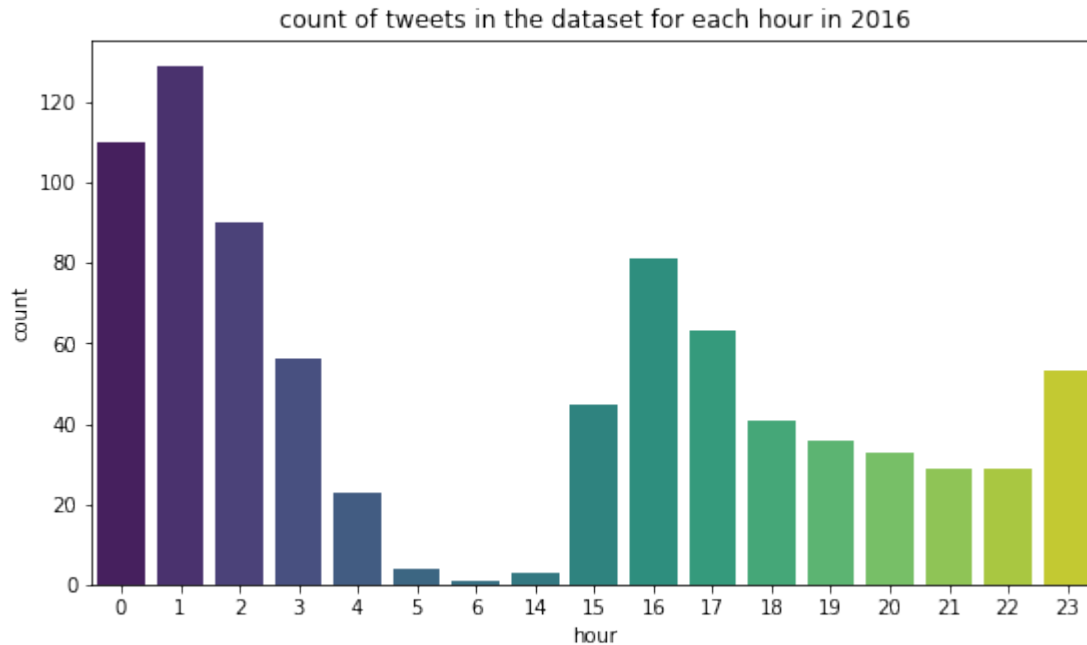
```
fig, ax = plt.subplots(figsize = (9,5))
ax = sns.countplot(x = "hour", data = df_time.query("year == 2015"), palette =
"viridis")
ax.set_title("count of tweets in the dataset for each hour in 2015");
```

count of tweets in the dataset for each hour in 2015
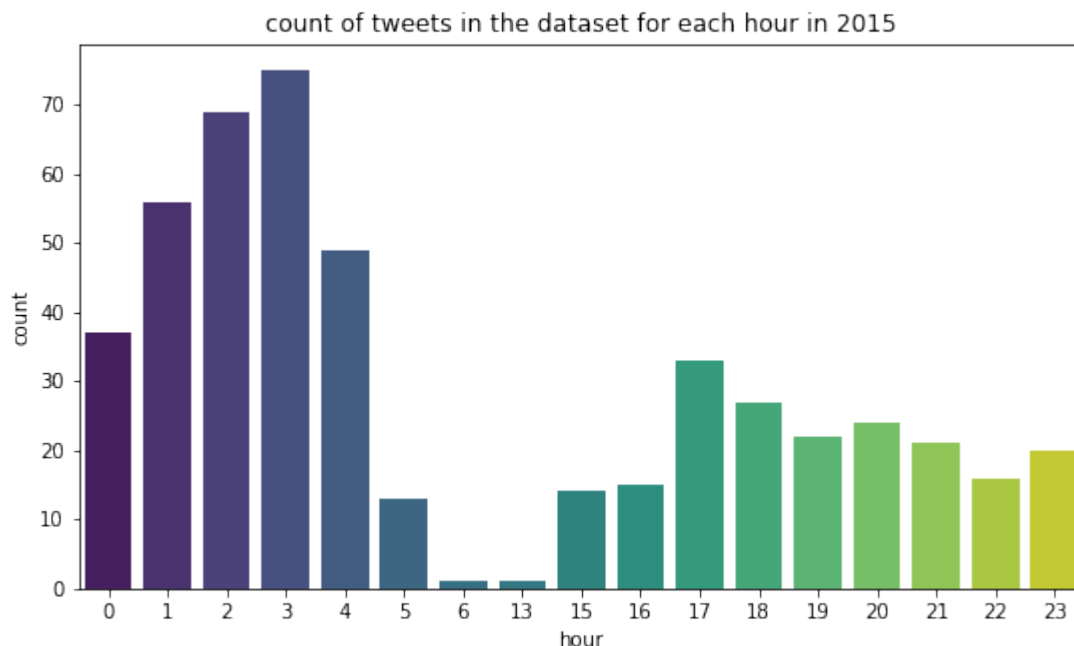
As we can see in this graphs of 2015 and 2016, the most posts are during the night between 0 (inlcuded)
and 5 'o clock. between 4 and 15 'o clock there is a very small amounts of tweets, in fact there is no single

tweet for the hours 7 - 12 and 14 o' clock. There are some tweets after 14 'o clock, but not as much as between 0 (included) and 5 'o clock.

Summary and Conclusions¶

In this project we went through a whole data wrangling process. The gathering included the programmatical, manual download as also the gathering via an API. Also the cleaning part covered a lot of techniques like regular expressions or the overall handling of dataframes.

In the analysis we took a look at following questions:

Questions:

1. Based on the predicted, most likely dog breed: Which breed gets retweeted and favorited the most overall?

   The winner for our analysis was the labrador retriever.

2. How did the account develop (speaking about number of tweets, retweets, favorites, image number and length of the tweets)?

We found, that the number of tweets per month decreased, while the retweets and favorites show an uptrend. For the image numbers there is no clear trend visible, the length of the tweets got a little bit closer to the maximum of 130 in the second half of the dataset.

1. Is there a pattern visible in the timing of the tweets?

Here we found, that between 5 and 15 'o clock there are nearly no tweets at all. The most tweets are during the time from 0 - 4 and then from 15 - 23, but from 15 - 23 are less tweets than between 0 - 4.

In [0]: